

Compressive Sensing over the Grassmann Manifold: a Unified Geometric Framework

Weiyu Xu and Babak Hassibi

Abstract

ℓ_1 minimization is often used for finding the sparse solutions of an under-determined linear system. In this paper we focus on finding sharp performance bounds on recovering approximately sparse signals using ℓ_1 minimization, possibly under noisy measurements. While the restricted isometry property is powerful for the analysis of recovering approximately sparse signals with noisy measurements, the known bounds on the achievable sparsity¹ level can be quite loose. The neighborly polytope analysis which yields sharp bounds for ideally sparse signals cannot be readily generalized to approximately sparse signals. Starting from a necessary and sufficient condition, the “balancedness” property of linear subspaces, for achieving a certain signal recovery accuracy, we give a unified *null space Grassmann angle*-based geometric framework for analyzing the performance of ℓ_1 minimization. By investigating the “balancedness” property, this unified framework characterizes sharp quantitative tradeoffs between the considered sparsity and the recovery accuracy of the ℓ_1 optimization. As a consequence, this generalizes the neighborly polytope result for ideally sparse signals. Besides the robustness in the “strong” sense for *all* sparse signals, we also discuss the notions of “weak” and “sectional” robustness. Our results concern fundamental properties of linear subspaces and so may be of independent mathematical interest.

I. INTRODUCTION

Compressive sensing is an area in signal processing which has attracted a lot of attention recently [Can06] [Don06a]. The motivation behind compressive sensing is to do “sampling” and

Weiyu Xu was with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, 91125 USA. e-mail: weiyu@systems.caltech.edu. He is now with School of Electrical and Computer Engineering, Cornell University.

Babak Hassibi is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, 91125 USA. e-mail: hassibi@systems.caltech.edu.

¹The “sparsity” in this paper means the size of the set of nonzero or significant elements in a signal vector.

“compression” at the same time. In conventional wisdom, in order to fully recover a signal, one has to sample the signal at a sampling rate equal or greater to the Nyquist sampling rate. The process of “sampling at full rate” and then “throwing away in compression” can prove to be wasteful of sensing and sampling resources, especially in application scenarios where resources like sensors, energy, observation time, etc. are limited. However, in many applications such as imaging, sensor networks, astronomy, biological systems [RIC], the signals of interest are often “sparse” over a certain basis. In these cases, compressive sensing promises to use a much smaller number of samples or measurements while still being able to recover the original sparse signal exactly or approximately. What enables practical compressive sensing is the existence of efficient decoding algorithms to recover the sparse signals from the “compressed” measurements. One of the most important and powerful decoding algorithms is the Basis Pursuit method, namely the ℓ_1 minimization method [CM73], [Don06b].

In this paper, we are interested in analyzing the decoding performance of the ℓ_1 minimization algorithm for approximately sparse signals under possibly noisy measurements. Mathematically, in compressive sensing problems, we would like to find an $n \times 1$ vector \mathbf{x} such that

$$\mathbf{y} = A\mathbf{x}, \tag{1}$$

where A is an $m \times n$ measurement matrix, \mathbf{y} is an $m \times 1$ measurement vector and $m < n$ in general. In the usual compressive sensing context \mathbf{x} is an $n \times 1$ unknown k -sparse vector, which has only k nonzero components. In this paper we will consider a more general version of the k -sparse vector \mathbf{x} . Namely, we will assume that k components of the vector \mathbf{x} have large magnitudes and that the vector comprised of the remaining $(n - k)$ components has an ℓ_1 -norm less than some value, say, Δ . We will refer to this type of signal as an approximately k -sparse signal, or for brevity only an approximately sparse signal. It is also possible that the \mathbf{y} can be further corrupted with measurement noise. This problem setup is more realistic of practical applications than the standard compressive sensing of ideally k -sparse signals (see, e.g., [TWD⁺06], [Can06], [CRT06] and the references therein). The interested readers can find more on similar type of problems in [CDD08] and other references.

In the rest of the paper we will further assume that the number of the measurements is $m = \delta n$ and the number of the “large” components of \mathbf{x} is $k = \rho \delta n = \zeta n$, where $0 < \rho < 1$ and $0 < \delta < 1$ are constants independent of n (clearly, $\delta > \zeta$).

A. ℓ_1 Minimization for Ideally Sparse Signal

ℓ_1 minimization optimization (Basis Pursuit) proposes solving the following problem

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_1 \\ \text{subject to} \quad & \mathbf{y} = A\mathbf{x}, \end{aligned} \tag{2}$$

where $\|\mathbf{x}\|_1$ denotes the ℓ_1 norm of \mathbf{x} , namely the sum of the amplitudes of all the elements in \mathbf{x} .

In [CT05] the authors were able to show that if the number of the measurements is $m = \delta n$ and if the matrix A satisfies a special property called the restricted isometry property (RIP), then any unknown vector \mathbf{x} with no more than $k = \zeta n$ (where ζ is an absolute constant as a function of δ , but independent of n , and explicitly bounded in [CT05]) nonzero elements can be recovered by solving (2). As expected, this assumes that \mathbf{y} was in fact generated by such an \mathbf{x} and given to us (more on the case when the available measurements are noisy versions of \mathbf{y} can be found in e.g. [HN], [Wai06]).

As can be immediately seen, the previous results heavily rely on the assumption that the measurement matrix A satisfies the RIP condition. It turns out that for several specific classes of matrices, such as matrices with independent zero-mean Gaussian entries or independent Bernoulli entries, the RIP holds with overwhelming probability [CT05], [BDDW08], [RV]. However, it should be noted that the RIP condition is only a sufficient condition for ℓ_1 -optimization to produce a solution of (1).

Instead of characterizing the $m \times n$ matrix A through the RIP condition, in [Don06b], [DT05], the authors proposed to study A through a k -neighborly polytope condition. As shown in [Don06b], this characterization of the matrix A is in fact a necessary and sufficient condition for (2) to produce the sparse solution \mathbf{x} satisfying (1). Furthermore, developing the results of [VS92], it can be shown that if the matrix A has i.i.d. zero-mean Gaussian entries, then the k -neighborly polytope condition holds with overwhelming probability. The precise relation between m , n and k in order for this to happen is characterized in [Don06b]. It should also be noted that for a given value m , i.e. for a given value of the constant δ , the value of the constant ζ given by the neighborly polytope condition is significantly better in [Don06b], [DT05] than in [CT05]. In fact, the values of ζ for the so-called “weak” threshold, obtained for different values of δ in [Don06b], approach the ones obtained by simulation as $n \rightarrow \infty$.

B. ℓ_1 Minimization for Approximately Sparse Signal

As mentioned earlier, in this paper we will be interested in recovering not perfectly k -sparse signals from compressed observations \mathbf{y} . In this case an exact recovery of the unknown vector \mathbf{x} from a reduced number of measurements is not possible in general. Instead, we will prove that, if we denote the unknown signal as \mathbf{x} , denote $\hat{\mathbf{x}}$ as one solution to (2), then for any given constant $0 < \delta < 1$ and any given constant $C > 1$ (representing how close in ℓ_1 norm the recovered vector $\hat{\mathbf{x}}$ should be to \mathbf{x}), there exists a constant $\zeta > 0$ and a sequence of measurement matrices $A \in \mathbb{R}^{m \times n}$ as $n \rightarrow \infty$ such that

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq \frac{2(C+1)\Delta}{C-1}, \quad (3)$$

holds for *all* $\mathbf{x} \in \mathbb{R}^n$, where Δ is the ℓ_1 norm of any $(n-k)$ elements of the vector \mathbf{x} (recall $k = \zeta n$). Here ζ will be a function of C and δ , but independent of the problem dimension n . In particular, we have the following theorem.

Theorem 1: Let $n, m, k, \mathbf{x}, \hat{\mathbf{x}}$ and Δ be defined as above. Let K denote a subset of $\{1, 2, \dots, n\}$ such that $|K| = k$, where $|K|$ is the cardinality of K , and let K_i denote the i -th element of K and $\overline{K} = \{1, 2, \dots, n\} \setminus K$.

Then for any constant $C > 1$ and any $\delta = \frac{m}{n} > 0$, there exists a $\zeta(\delta, C) > 0$ such that if the measurement matrix A is the basis for a uniformly-distributed subspace, then with overwhelming probability as $n \rightarrow \infty$, for all vectors $\mathbf{w} \in \mathbb{R}^n$ in the null space of A , and for all K such that $|K| = k \leq \zeta(\delta, C)n$, we have

$$C \sum_{i=1}^k |\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k} |\mathbf{w}_{\overline{K}_i}|, \quad (4)$$

where \mathbf{x}_K denotes the part of \mathbf{x} over the subset K ; and at the same time the solution $\hat{\mathbf{x}}$ produced by (2) will satisfy

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq \frac{2(C+1)\Delta}{C-1}. \quad (5)$$

for *all* $\mathbf{x} \in \mathbb{R}^n$.

The main focus and contribution of this paper is to establish a sharp relationship between δ , ζ and C . For example, when $\delta = \frac{m}{n}$ varies, we have Figure 1 showing the tradeoff between the signal sparsity ζ and the parameter C , which determines the robustness² of the ℓ_1 minimization.

²The “robustness” concept in this sense is often called the “stability” in other papers, for example, [Can06].

The curve for $C = 1$ matches the “strong” threshold curve from [Don06b] for ideally sparse signal vectors .

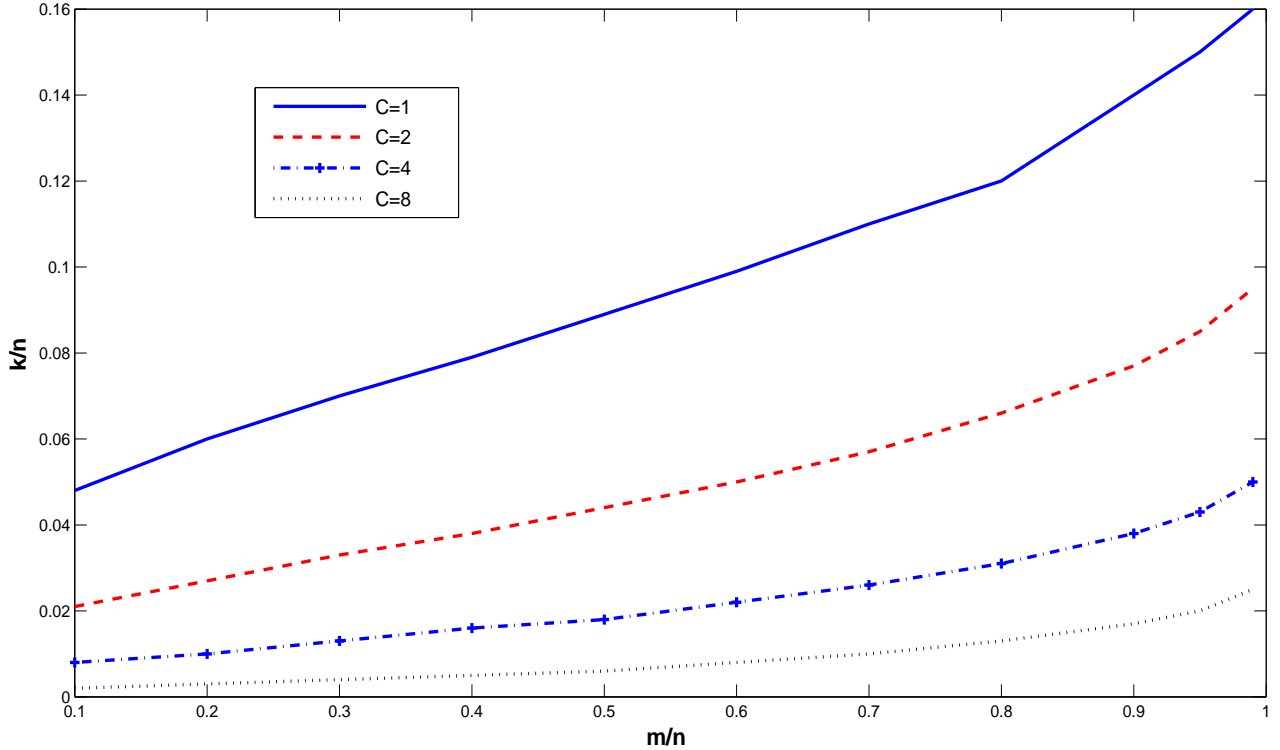


Fig. 1: Tradeoff between signal sparsity and ℓ_1 recovery robustness as a function of C (allowable imperfection of the recovered signal is $\frac{2(C+1)\Delta}{C-1}$)

To obtain the stated results we will make use of a characterization that constitutes both necessary and sufficient conditions on the matrix A such that the solution of (2) approximates the original signal accurately enough such that (3) holds. This characterization will be equivalent to the neighborly polytope characterization from [Don06b] in the “ideally sparse” case when $C = 1$. Furthermore, as we will see later in the paper, in the perfectly sparse signal case (which allows $C \rightarrow 1$), our result for allowable ζ matches the result of [Don06b]. Our analysis will be directly based on the null space Grassmann angle approach in high dimensional integral geometry, which gives a unified analytical framework for ℓ_1 minimization.

A similar problem was considered in [CDD08], where the null space characterization for recovering approximately sparse signal was analyzed using the RIP in [CT05]; however, no explicit values of ζ were given. Since the RIP condition is a sufficient condition for good sparse signal recoveries using ℓ_1 minimization, it generally gives rather loose bounds on the explicit values of ζ even in the ideally sparse signal case [CT05][BCT09]. There have also been some recent works trying to analyze the performance of ℓ_1 minimization through non-RIP techniques [Zha08], [Vav09], [Sto09]. Compared with previous results, in this paper we will provide sharp bounds on the explicit values of the allowable constants ζ for satisfying the subspace “balancedness” condition, as a function of $C \geq 1$. In the literature, there are also discussions of compressive sensing under different definitions of non-ideally sparse signals, for example, [Don06a] discusses compressive sensing for signals from a ℓ_p ball with $0 < p \leq 1$ using sufficient conditions based on results of the Gelfand n -widths. However, the results of this paper are dealing directly with approximately sparse signals defined in terms of the concentration of ℓ_1 norm, and furthermore, we give a neat *necessary and sufficient* condition for ℓ_1 optimization to be robust and we are also able to explicitly give much sharper bounds on the sparsity parameter ζ . When we finalize this draft from our earlier conference publication [XH08], we are informed of the very recent work [DMM10] which deals with a related but different problem formulation of characterizing the tradeoff between signal sparsity and noise sensitivity of LASSO recovery method. Compared with [DMM10], we are dealing with the plain ℓ_1 minimization method for recovering approximately sparse signals, and the performance bounds in this paper apply to general type of signals and noises. The analysis from [DMM10] is an average-case analysis for compressed measurements corrupted with Gaussian noises, while the analysis in this paper provides both average-case and worst-case performance bounds under general types of signals and noises. It is also noteworthy pointing out that this work considers the plain ℓ_1 minimization, which does not require the decoder to know of the statistical variance of the measurement noises. The analysis methodologies between this work and [DMM10] are also different: this work relies on the analytical tools from the high dimensional polytope geometry, while [DMM10] builds on the innovations of analyzing message passing algorithms.

The rest of the paper is organized as follows. In Section II, we introduce a null space characterization of linear subspaces for guaranteeing robust signal recovery using the ℓ_1 minimization. Section III presents a Grassmann angle-based high dimensional geometrical framework

for analyzing the null space characterization. In Sections IV, VI, and VII, analytical performance bounds are given for the null space characterization. Section VIII shows how the Grassmann angle analytical framework can be extended to analyzing the “weak”, “sectional” and “strong” notations of robust signal recovery. In Section IX, we present the robustness analysis of the ℓ_1 minimization under noisy measurements. In Section X, the numerical evaluations of the performance bounds for robust signal recovery are given. Section XI concludes the paper. In the appendix, we provide a quick summary of relevant geometric concepts in the high dimensional geometry and the proofs of related lemmas and theorems.

II. THE NULL SPACE CHARACTERIZATION

In this section we introduce a useful characterization of the matrix A . The characterization will establish a necessary and sufficient condition on the matrix A so that the solution of (2) approximates the solution of (1) such that (3) holds. (See [FN03], [LN06], [Zha06], [CDD08], [SXH08a], [SXH08b], [KT07] etc. for variations of this result).

Theorem 2: Assume that A is a general $m \times n$ measurement matrix. Let $C > 1$ be a positive number. Further, assume that $\mathbf{y} = A\mathbf{x}$ and that \mathbf{w} is an $n \times 1$ vector. Let K be a subset of $\{1, 2, \dots, n\}$ such that $|K| = k$, where $|K|$ is the cardinality of K and let K_i denote the i -th element of K . Further, let $\overline{K} = \{1, 2, \dots, n\} \setminus K$. Then for any $\mathbf{x} \in \mathbb{R}^n$, for any K such that $|K| = k$, any solution $\hat{\mathbf{x}}$ produced by (2) will satisfy

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1} \|\mathbf{x}_{\overline{K}}\|_1, \quad (6)$$

if $\forall \mathbf{w} \in \mathbb{R}^n$ such that

$$A\mathbf{w} = 0$$

and $\forall K$ such that $|K| = k$, we have

$$C \sum_{i=1}^k |\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k} |\mathbf{w}_{\overline{K}_i}|. \quad (7)$$

Conversely, there exists some measurement matrix A , a set K with cardinality k , an \mathbf{x} , and corresponding $\hat{\mathbf{x}}$ ($\hat{\mathbf{x}}$ is a minimizer to the programming (2)), such that (7) is satisfied with equality for some vector \mathbf{w} in the null space of A with a constant $C' > 1$; moreover

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 = 2 \frac{(C'+1)}{C'-1} \|\mathbf{x}_{\overline{K}}\|_1.$$

and

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 > 2 \frac{(C+1)}{C-1} \|\mathbf{x}_{\bar{K}}\|_1.$$

for any C bigger than the constant C' .

Proof: First, suppose the matrix A has the claimed null space property as in (7) and we want to prove that any solution $\hat{\mathbf{x}}$ satisfies (6). Note that the solution $\hat{\mathbf{x}}$ of (2) satisfies

$$\|\hat{\mathbf{x}}\|_1 \leq \|\mathbf{x}\|_1,$$

where \mathbf{x} is the original signal. Since $A\hat{\mathbf{x}} = \mathbf{y}$, it easily follows that $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$ is in the null space of A . Therefore we can further write $\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{w}\|_1$. Using the triangular inequality for the ℓ_1 norm we obtain

$$\begin{aligned} \|\mathbf{x}_K\|_1 + \|\mathbf{x}_{\bar{K}}\|_1 &= \|\mathbf{x}\|_1 \\ &\geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x} + \mathbf{w}\|_1 \\ &\geq \|\mathbf{x}_K\|_1 - \|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\bar{K}}\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\ &\geq \|\mathbf{x}_K\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 + \frac{C-1}{C+1} \|\mathbf{w}\|_1 \end{aligned}$$

where the last inequality is from the claimed null space property. Relating the head and tail of the inequality chain above,

$$2\|\mathbf{x}_{\bar{K}}\|_1 \geq \frac{(C-1)}{C+1} \|\mathbf{w}\|_1.$$

Now we prove the second part of the theorem, namely when (7) is violated, there exist scenarios where the error performance bound (6) fails. The simplest example is when the null space of the measurement matrix A is a one-dimensional subspace and has an all-1 vector $(1, 1, \dots, 1)$ as its basis. Let n be an even number. For any $k < \frac{n}{2}$, let us take $C' = \frac{n-k}{k}$ and $C = \frac{n-k}{k} + \epsilon$, where $\epsilon > 0$ is an arbitrarily small positive number. Then obviously there exists a vector \mathbf{w} in the null space of A that violates the condition (7) for $C = \frac{n-k}{k} + \epsilon$ for the set $K = \{1, 2, \dots, k\}$. Now we consider a signal vector

$$\mathbf{x} = (\underbrace{-1, -1, \dots, -1}_{\frac{n}{2}}, \underbrace{0, 0, \dots, 0}_{\frac{n}{2}}).$$

Taking the null space of A into account, we can see

$$\hat{\mathbf{x}} = (\underbrace{0, 0, \dots, 0}_{\frac{n}{2}}, \underbrace{1, 1, \dots, 1}_{\frac{n}{2}})$$

is a minimizer to the programming (2).

Note that $\|\mathbf{x}_{\bar{K}}\|_1 = \frac{n}{2} - k$ and $\|\mathbf{x} - \hat{\mathbf{x}}\|_1 = n$,

$$\begin{aligned} & \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_1}{\|\mathbf{x}_{\bar{K}}\|_1} \\ &= \frac{n}{\frac{n}{2} - k} = \frac{2(\frac{n-k}{k} + 1)}{\frac{n-k}{k} - 1} \\ &= \frac{2(C' + 1)}{(C' - 1)} \\ &> \frac{2(C + 1)}{C - 1}, \end{aligned}$$

strictly contradicting the error bound (6). ■

It should be noted that if the condition (7) is true for all the sets K of cardinality k , then

$$2\|\mathbf{x}_{\bar{K}}\|_1 \geq \frac{(C - 1)}{C + 1} \|\hat{\mathbf{x}} - \mathbf{x}\|_1$$

is also true for the set K which corresponds to the k largest (in amplitude) components of the vector \mathbf{x} . So

$$2\Delta \geq \frac{(C - 1)}{C + 1} \|\hat{\mathbf{x}} - \mathbf{x}\|_1$$

which exactly corresponds to (3). In fact, the condition (7) is also a sufficient and necessary condition for unique exact recovery of ideally k -sparse signals after we take $C = 1$ and let (7) take strict inequality for all $\mathbf{w} \neq 0$ in the null space of A . To see this, suppose the ideally k -sparse signal \mathbf{x} is supported over the set K , namely, $\|\mathbf{x}_{\bar{K}}\|_1 = 0$. Then from the same triangular inequality derivation of Theorem 2, we know that $\|\hat{\mathbf{x}} - \mathbf{x}\|_1 = 0$, namely $\hat{\mathbf{x}} = \mathbf{x}$. Or we can just let C be arbitrarily close to 1 from the right and since

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C + 1)}{C - 1} \|\mathbf{x}_{\bar{K}}\|_1 = 0,$$

we also get $\hat{\mathbf{x}} = \mathbf{x}$. In this sense, when $C = 1$, the null space condition is equivalent to the neighborly polytope condition [Don06b] for unique exact recovery of ideally sparse signals.

However, it is an interesting result that, for a particular *fixed* measurement matrix A , the violation of (7) for some $C > 1$ does not necessarily mean that the existence of a vector \mathbf{x} and

a minimizer solution $\hat{\mathbf{x}}$ to (2) such that the performance guarantee (6) is violated. For example, assume $n = 2$ and the null space of the measurement matrix A is a one-dimensional subspace and has the vector $(1, 100)$ as its basis. Then the null space of the matrix A violates (7) with $C = 101$ and the set $K = \{1\}$. But a careful examination shows that the biggest possible $\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_1}{\|\mathbf{x}_{\overline{K}}\|_1}$ ($\|\mathbf{x}_{\overline{K}}\|_1 \neq 0$) is equal to $\frac{100+1}{100} = \frac{101}{100}$, achieved by such an \mathbf{x} as $(-1, -1)$. In fact, all those vectors $\mathbf{x} = (a, b)$ with $b \neq 0$ will achieve $\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_1}{\|\mathbf{x}_{\overline{K}}\|_1} = \frac{101}{100}$. However, (6) has $\frac{2(C+1)}{C-1} = \frac{204}{100}$. This suggests that for a specific measurement matrix A , the tightest error bound for $\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_1}{\|\mathbf{x}_{\overline{K}}\|_1}$ should involve the detailed structure of the null space of A . But for general measurement matrices A , as suggested by Theorem 2, the condition (7) is a necessary and sufficient condition to offer the performance guarantee (6).

It is worth pointing out that the example given in the proof of Theorem 2 is not just an isolated example. In fact, for two general positive integers m and n with $m < n$ and $n \geq 2$, we can often find an $m \times n$ measurement matrix A and a certain $C > 1$ such that the condition (7) is violated and, at the same time, for some vector \mathbf{x} , the performance bound is also “tightly” violated.

Consider a generic $m \times n$ matrix A' . For each integer $1 \leq k \leq n$, let us define the quantity h_k as the supremum of $\frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}_{\overline{K}}\|_1}$ over all such sets K of size $|K| \leq k$ and over all nonzero vectors \mathbf{w} in the null space of A' . Let k^* be the biggest k such that $h_k \leq 1$. Then there must be a nonzero vector \mathbf{w}' in the null space of A and a set K^* of size k^* , such that

$$\|\mathbf{w}'_{K^*}\|_1 = h_{k^*} \|\mathbf{w}'_{\overline{K^*}}\|_1.$$

Now we generate a new measurement matrix A by multiplying the portion A'_{K^*} of the matrix A' by h_{k^*} . Then we will have a vector \mathbf{w} in the null space of A satisfying

$$\|\mathbf{w}_{K^*}\|_1 = \|\mathbf{w}_{\overline{K^*}}\|_1.$$

Now we take a signal vector $\mathbf{x} = (-\mathbf{w}_{K^*}, 0_{\overline{K^*}})$ and claim that $\hat{\mathbf{x}} = (0, \mathbf{w}_{\overline{K^*}})$ is a minimizer to the programming (2). In fact, recognizing the definition of h_{k^*} , we know all the vectors \mathbf{w}'' in the null space of the measurement matrix A will satisfy $\|\mathbf{x} + \mathbf{w}''\|_1 \geq \|\mathbf{x}\|_1$. Let us assume that $k^* \geq 2$ and take $K'' \subseteq K^*$ as the index set corresponding to the largest $(k^* - i)$ elements of \mathbf{x}_{K^*} in amplitude, where $1 \leq i \leq (k^* - 1)$. From the definition of k^* , it is apparent that $C' = \frac{\|\mathbf{w}_{K''}\|_1}{\|\mathbf{w}_{K''}\|_1} > 1$ since \mathbf{w} is nonzero for any index in the set K^* . Let us now take $C = \frac{\|\mathbf{w}_{K''}\|_1}{\|\mathbf{w}_{K''}\|_1} + \epsilon$,

where $\epsilon > 0$ is any arbitrarily small positive number. Thus the condition (7) is violated for the vector \mathbf{w} , the set K'' and the defined constant C .

Now by inspection, the decoding error is

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 = \frac{2(C' + 1)}{C' - 1} \|\mathbf{x}_{K''}\|_1 > \frac{2(C + 1)}{C - 1} \|\mathbf{x}_{K''}\|_1,$$

violating the error bound (6) (for the set K'').

In the remaining part of this paper, for a given value $\delta = \frac{m}{n}$ and any value $C \geq 1$, we will devote our efforts to determining the value of feasible $\zeta = \rho\delta = \frac{k}{n}$ for which there exists a sequence of A such that the null space condition (7) is satisfied for all the sets K of size k when n goes to infinity and $\frac{m}{n} = \delta$. For a specific A , it is very hard to check whether the condition (7) is satisfied or not. Instead, we consider randomly choosing A from a Gaussian distribution, and analyze for what ζ , the condition (7) for its null space is satisfied with overwhelming probability as n goes to infinity. When we consider $C = 1$, corresponding to the success of ℓ_1 minimization for all ideally k -sparse signals, loose bounds on ζ achieving the null space condition were established in [CT05][Zha06][SXH08a] using the restricted isometry property and high dimensional geometrical results. The null space condition is equivalent to the k -neighborly polytope condition when $C = 1$, so the neighborly polytope condition [Don06b] gives much sharper bounds for the null space condition when $C = 1$. However, no sharp bounds are available for the null space condition with the general case $C \geq 1$.

The standard results on compressive sensing assume that the matrix A has i.i.d. $\mathcal{N}(0, 1)$ entries. The following lemma gives a characterization of the resulting null space of A , which is a fairly well known result, and for the sake of completeness, we include its proof in the appendix.

Lemma 1: Let $A \in \mathbb{R}^{m \times n}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. Then the following statements hold:

- The distribution of A is right-rotationally invariant: for any Θ satisfying $\Theta\Theta^* = \Theta^*\Theta = I$, $P_A(A) = P_A(A\Theta)$;
- There exists a basis Z of the null space of A , such that the distribution of Z is left-rotationally invariant: for any Θ satisfying $\Theta\Theta^* = \Theta^*\Theta = I$, $P_Z(Z) = P_Z(\Theta^*Z)$;
- It is always possible to choose a basis Z for the null space such that Z has i.i.d. $\mathcal{N}(0, 1)$ entries.

In view of Theorem 1 and Lemma 1 what matters is that the null space of A be rotationally invariantly. Sampling from this rotationally invariant distribution is equivalent to uniformly sampling a random $(n - m)$ -dimensional subspace from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$. Here the Grassmann manifold $\text{Gr}_{(n-m)}(n)$ is the set of $(n - m)$ -dimensional subspaces in the n -dimensional Euclidean space \mathbb{R}^n [Boo86]. For any such A and ideally sparse signals, the sharp bounds of [Don06b], apply. However, we shall see that the neighborly polytope condition for ideally sparse signals does not readily apply to the proposed null space condition analysis for approximately sparse signals, since the null space condition can not be transformed to the k -neighborly property in a *single* high-dimensional polytope [Don06b]. Instead, in this paper, we shall give a unified Grassmann angle framework to analyze the proposed null space property.

III. THE GRASSMANN ANGLE FRAMEWORK FOR THE NULL SPACE CHARACTERIZATION

In this section we detail the Grassmann angle-based framework for analyzing the bounds on $\zeta = \frac{k}{n}$ such that (7) holds for every vector in the null space, which we denote by Z . Put more precisely, given a certain constant $C > 1$ (or $C \geq 1$), which corresponds to a certain level of recovery accuracy for the approximately sparse signals, we are interested in what scaling $\frac{k}{n}$ we can achieve while satisfying the following condition on Z ($|K| = k$):

$$\forall \mathbf{w} \in Z, \forall K \subseteq \{1, 2, \dots, n\}, C \|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\overline{K}}\|_1. \quad (8)$$

From the definition of the condition (8), there is a tradeoff between the largest sparsity level k and the parameter C . As C grows, clearly the largest k satisfying (8) will likely decrease, and, at the same time, ℓ_1 minimization will be more robust in terms of the the residual norm $\|\mathbf{x}_{\overline{K}}\|_1$. The key in our derivation is the following lemma:

Lemma 2: For a certain subset $K \subseteq \{1, 2, \dots, n\}$ with $|K| = k$, the event that the null space Z satisfies

$$C \|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\overline{K}}\|_1, \forall \mathbf{w} \in Z$$

is equivalent to the event that $\forall \mathbf{x}$ supported on the k -set K (or supported on a subset of K):

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1 \geq \|\mathbf{x}_K\|_1, \forall \mathbf{w} \in Z. \quad (9)$$

Proof: First, let us assume that $C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\overline{K}}\|_1, \forall \mathbf{w} \in Z$. Using the triangular inequality, we obtain

$$\begin{aligned} & \|\mathbf{x}_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1 \\ & \geq \|\mathbf{x}_K\|_1 - \|\mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1 \\ & \geq \|\mathbf{x}_K\|_1. \end{aligned}$$

thus proving the forward part of this lemma. Now let us assume instead that $\exists \mathbf{w} \in Z$, such that $C\|\mathbf{w}_K\|_1 > \|\mathbf{w}_{\overline{K}}\|_1$. Then we can construct a vector \mathbf{x} supported on the set K (or a subset of K), with $\mathbf{x}_K = -\mathbf{w}_K$. Then we have

$$\begin{aligned} & \|\mathbf{x}_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1 \\ & = 0 + \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1 \\ & < \|\mathbf{x}_K\|_1, \end{aligned}$$

proving the inverse part of this lemma. ■

Now let us consider the probability that condition (8) holds for the sparsity $|K| = k$ if we uniformly sample a random $(n - m)$ -dimensional subspace Z from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$. Based on Lemma 2, we can equivalently consider the complementary probability P that there exists a subset $K \subseteq \{1, 2, \dots, n\}$ with $|K| = k$, and a vector $\mathbf{x} \in \mathbb{R}^n$ supported on the set K (or a subset of K) failing the condition (9). With the linearity of the subspace Z in mind, to obtain P , we can restrict our attention to those vectors \mathbf{x} from the cross-polytope (the unit ℓ_1 ball)

$$\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_1 = 1\}$$

that are only supported on the set K (or a subset of K).

First, we upper bound the probability P by a union bound over all the possible support sets $K \subseteq \{1, 2, \dots, n\}$ and all the sign patterns of the k -sparse vector \mathbf{x} . Since the k -sparse vector \mathbf{x} has $\binom{n}{k}$ possible support sets of cardinality k and 2^k possible sign patterns (nonnegative or nonpositive), we have

$$P \leq \binom{n}{k} \times 2^k \times P_{K,-}, \quad (10)$$

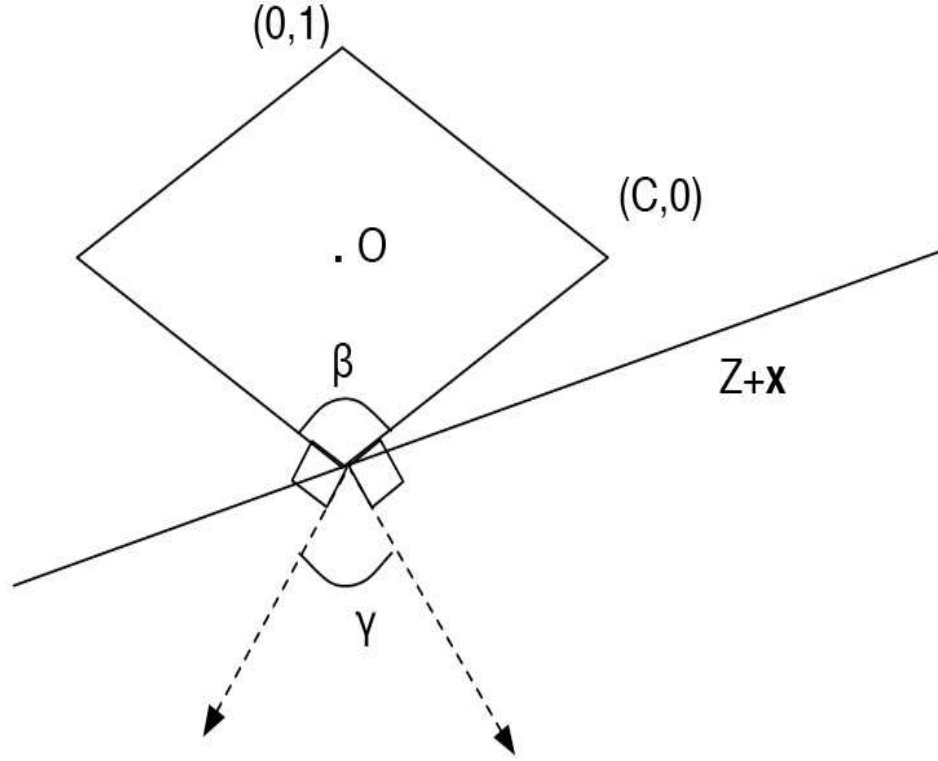


Fig. 2: The Grassmann Angle for a Skewed Cross-polytope

where $P_{K,-}$ is the probability that for a specific *support set* K , there exist a k -sparse vector \mathbf{x} of a specific *sign pattern* which fails the condition (9). By symmetry, without loss of generality, we assume the signs of the elements of \mathbf{x} to be nonpositive.

So now let us focus on deriving the probability $P_{K,-}$. Since \mathbf{x} is a nonpositive k -sparse vector supported on the set K (or a subset of K) and can be restricted to the cross-polytope $\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_1 = 1\}$, \mathbf{x} is also on a $(k-1)$ -dimensional face, denoted by F , of the skewed cross-polytope (weighted ℓ_1 ball) SP:

$$\text{SP} = \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\} \quad (11)$$

Then $P_{K,-}$ is the probability that there exists an $\mathbf{x} \in F$, and there exists a $\mathbf{w} \in Z$ ($\mathbf{w} \neq 0$) such that

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \leq \|\mathbf{x}_K\|_1 = 1. \quad (12)$$

We first focus on studying a specific *single* point $\mathbf{x} \in F$, without loss of generality, assumed to be in the relative interior of this $(k-1)$ dimensional face F . For this single particular \mathbf{x} on the F , the probability, denoted by P'_x , that $\exists \mathbf{w} \in Z$ ($\mathbf{w} \neq 0$) such that 12 holds is essentially the probability that a uniformly chosen $(n-m)$ dimensional subspace Z shifted by the point \mathbf{x} , namely $(Z + \mathbf{x})$, intersects the skewed cross-polytope

$$\text{SP} = \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\} \quad (13)$$

nontrivially, namely, at some other point besides \mathbf{x} .

From the linear property of the subspace Z , the event that $(Z + \mathbf{x})$ intersects the skewed cross-polytope SP is equivalent to the event that Z intersects nontrivially with the cone $\text{SP-Cone}(\mathbf{x})$ obtained by observing the skewed polytope SP from the point \mathbf{x} . (Namely, $\text{SP-Cone}(\mathbf{x})$ is conic hull of the point set $(\text{SP} - \mathbf{x})$ and $\text{SP-Cone}(\mathbf{x})$ has the origin of the coordinate system as its apex.) However, as noticed in the geometry for convex polytopes [Grü68][Grü03], the $\text{SP-Cone}(\mathbf{x})$ are identical for any \mathbf{x} lying in the relative interior of the face F . This means that the probability $P_{K,-}$ is equal to P'_x , regardless of the fact \mathbf{x} is only a single point in the relative interior of the face F . (The acute reader may have noticed some singularities here because $\mathbf{x} \in F$ may not be in the relative interior of F , but it turns out that the $\text{SP-Cone}(\mathbf{x})$ is then only a subset of the cone we get when \mathbf{x} is in the relative interior of F . So we do not lose anything if we restrict \mathbf{x} to be in the relative interior of the face F .) In summary, we have

$$P_{K,-} = P'_x.$$

Now we only need to determine P'_x . From its definition, P'_x is exactly the **complementary Grassmann angle** [Grü68] for the face F with respect to the polytope SP under the Grassmann manifold $\text{Gr}_{(n-m)}(n)$:³ the probability of a uniformly distributed $(n-m)$ -dimensional subspace Z from the Grassmannian manifold $\text{Gr}_{(n-m)}(n)$ intersecting nontrivially with the cone $\text{SP-Cone}(\mathbf{x})$ formed by observing the skewed cross-polytope SP from the relative interior point $\mathbf{x} \in F$.

Building on the works by L.A.Santaló [San52] and P.McMullen [McM75] etc. in high dimensional integral geometry and convex polytopes, the complementary Grassmann angle for the

³A Grassman angle and its corresponding complementary Grassmann angle always sum up to 1. There is apparently inconsistency in terms of the definition of which is “Grassmann angle” and which is “complementary Grassmann angle” between [Grü68],[AS92] and [VS92] etc. But we will stick to the earliest definition in [Grü68] for Grassmann angle: the measure of the subspaces that intersect trivially with a cone.

$(k - 1)$ -dimensional face F can be explicitly expressed as the sum of products of internal angles and external angles [Grü03]:

$$2 \times \sum_{s \geq 0} \sum_{G \in \mathfrak{S}_{m+1+2s}(\text{SP})} \beta(F, G) \gamma(G, \text{SP}), \quad (14)$$

where s is any nonnegative integer, G is any $(m + 1 + 2s)$ -dimensional face of the skewed cross-polytope ($\mathfrak{S}_{m+1+2s}(\text{SP})$ is the set of all such faces), $\beta(\cdot, \cdot)$ stands for the internal angle and $\gamma(\cdot, \cdot)$ stands for the external angle.

The internal angles and external angles are basically defined as follows [Grü03][McM75]:

- An internal angle $\beta(F_1, F_2)$ is the fraction of the hypersphere S covered by the cone obtained by observing the face F_2 from the face F_1 .⁴ The internal angle $\beta(F_1, F_2)$ is defined to be zero when $F_1 \not\subseteq F_2$ and is defined to be one if $F_1 = F_2$.
- An external angle $\gamma(F_3, F_4)$ is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face F_4 at the face F_3 . The external angle $\gamma(F_3, F_4)$ is defined to be zero when $F_3 \not\subseteq F_4$ and is defined to be one if $F_3 = F_4$.

Let us take for example the 2-dimensional skewed cross-polytope

$$\text{SP} = \{(y_1, y_2) \in \mathbb{R}^2 \mid \|y_2\|_1 + \|\frac{y_1}{C}\|_1 \leq 1\}$$

(namely the diamond) in Figure 2, where $n=2$, $(n - m) = 1$ and $k = 1$. Then the point $\mathbf{x} = (0, -1)$ is a 0-dimensional face (namely a vertex) of the skewed polytope SP. Now from their definitions, the internal angle $\beta(\mathbf{x}, \text{SP}) = \frac{\beta}{2\pi}$ and the external angle $\gamma(\mathbf{x}, \text{SP}) = \frac{\gamma}{2\pi}$, $\gamma(\text{SP}, \text{SP}) = 1$. The complementary Grassmann angle for the vertex \mathbf{x} with respect to the polytope SP is the probability that a uniformly sampled 1-dimensional subspace (namely a line, we denote it by Z) shifted by \mathbf{x} intersects nontrivially with $\text{SP} = \{(y_1, y_2) \in \mathbb{R}^2 \mid \|y_2\|_1 + \|\frac{y_1}{C}\|_1 \leq 1\}$ (or equivalently the probability that Z intersects nontrivially with the cone obtained by observing SP from the point \mathbf{x}). It is obvious that this probability is $\frac{\beta}{\pi}$. The readers can also verify the correctness of the formula (14) very easily for this toy example.

Generally, it might be hard to give explicit formulae for the external and internal angles involved, but fortunately in the skewed cross-polytope case, both the internal angles and the external angles can be explicitly computed.

⁴Note the dimension of the hypersphere S here matches the dimension of the corresponding cone discussed. Also, the center of the hypersphere is the apex of the corresponding cone. All these defaults also apply to the definition of the external angles.

Firstly, let us look at the internal angle $\beta(F, G)$ between the $(k - 1)$ -dimensional face F and a $(l - 1)$ -dimensional face G . Notice that the only interesting case is when $F \subseteq G$ since $\beta(F, G) \neq 0$ only if $F \subseteq G$. We will see if $F \subseteq G$, the cone formed by observing G from F is the direct sum of a $(k - 1)$ -dimensional linear subspace and a convex polyhedral cone formed by $(l - k)$ unit vectors with inner product $\frac{1}{1+C^2k}$ between each other. In this case, the internal angle is given by

$$\beta(F, G) = \frac{V_{l-k-1}(\frac{1}{1+C^2k}, l-k-1)}{V_{l-k-1}(S^{l-k-1})}, \quad (15)$$

where $V_i(S^i)$ denotes the i -th dimensional surface measure on the unit sphere S^i , while $V_i(\alpha', i)$ denotes the surface measure for regular spherical simplex with $(i + 1)$ vertices on the unit sphere S^i and with inner product as α' between these $(i + 1)$ vertices. Thus (15) is equal to $B(\frac{1}{1+C^2k}, l-k)$, where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m' - 1)\alpha' + 1} \pi^{-m'/2} \alpha'^{-1/2} J(m', \theta), \quad (16)$$

with $\theta = (1 - \alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\theta v^2 + 2iv\lambda} dv \right)^{m'} e^{-\lambda^2} d\lambda. \quad (17)$$

We should remark that the formula above for the internal angle is true only when the face G is not of dimension n . When G is n -dimensional, we will derive a separate formula in Lemma 15. Since the expression for this special case will not affect our following derivations in a significant way, we choose not to list it here.

Secondly, we can derive the external angle $\gamma(G, \text{SP})$ between the $(l - 1)$ -dimensional face G and the skewed cross-polytope SP as:

$$\gamma(G, \text{SP}) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^{\infty} e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx. \quad (18)$$

The derivations of these expressions involve the computations of the volumes of cones in high dimensional geometry and will be presented in the appendix.

In summary, combining (10), (14), (15) and (18), we get an upper bound on the probability P . If we can show that for a certain $\zeta = \frac{k}{n}$, P goes to zero exponentially in n as $n \rightarrow \infty$, then we know that for such ζ , the null space condition (8) holds with overwhelming probability. This is the guideline for computing the bound on ζ in the following sections.

IV. EVALUATING THE BOUND ζ

In summary,

$$P \leq \binom{n}{k} \times 2^k \times 2 \times \sum_{s \geq 0} \sum_{G \in \mathfrak{S}_{m+1+2s}(\text{SP})} \beta(F, G) \gamma(G, \text{SP}). \quad (19)$$

In order for this upper bound on P to decrease to 0 as $n \rightarrow \infty$, one sufficient condition is that every sum term in (19) goes to 0 exponentially fast in n . We remark that the equation in (19) is similar to the expected number of missed “faces” in the study of k -neighborly polytope [Don06b], [VS92], but generalizes the k -neighborly polytope formula to more general Grassmann angles. In the following sections, we will extend the techniques developed in [Don06b], [VS92] to evaluating the bounds on ζ from (19), taking into account of the variable $C > 1$. To illustrate the effect of C on the bound ζ , also for the sake of completeness, we will keep the detailed derivations.

For simplicity of analysis, we define $l = (m + 1 + 2s) + 1$ and $\nu = \frac{l}{n}$. In the skewed cross-polytope SP , we notice that there are in total $\binom{n-k}{l-k} 2^{l-k}$ faces G of dimension $(l-1)$ such that $F \subseteq G$ and $\beta(F, G) \neq 0$. Because of the symmetry in the skewed cross-polytope SP , it follows from (19) that

$$P \leq \sum_{s \geq 0} \underbrace{2 \binom{n}{k} 2^l \times \binom{n-k}{l-k}}_{\substack{COM_s \\ D_s}} \beta(F, G) \gamma(G, \text{SP}), \quad (20)$$

where $l = (m + 1 + 2s) + 1$ and $G \subseteq \text{SP}$ is any single face of dimension $(l-1)$ such that $F \subseteq G$.

Closely following the approach of [Don06b], in estimating $n^{-1} \log(D_s)$, we can decompose it into a sum of terms involving logarithms of the combinatorial factor, the internal angle and the external angle. With

$$H(p) = p \log(1/p) + (1-p) \log(1/(1-p)),$$

where the logarithm base is over e . From Stirling's formula, we know that

$$n^{-1} \log \binom{n}{\lfloor pn \rfloor} \rightarrow H(p), p \in [0, 1], n \rightarrow \infty. \quad (21)$$

Defining $\nu = l/n \geq \delta$, we have

$$n^{-1} \log(COM_s) = \nu \log(2) + H(\rho\delta) + H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta) + R_1 \quad (22)$$

with remainder $R_1 = R_1(s, k, m, n)$.

Define the combinatorial growth exponent for COM_s

$$\psi_{com}(\nu; \rho, \delta) = \nu \log(2) + H(\rho\delta) + H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta), \quad (23)$$

describing the exponential growth of the combinatorial factors. Applying (21), we will see that the remainder R_1 in (22) is $o(1)$ uniformly in the range $l > \delta n$, $n > n_0(\rho, \delta, \epsilon)$, where $n_0(\rho, \delta, \epsilon)$ is some big enough natural number.

For a particular C , we will also define a decay exponent $\psi_{ext}(\nu; \rho, \delta)$ and show that $\gamma(G, SP)$ decays exponentially at least at the rate $\psi_{ext}(\nu; \rho, \delta)$: for each $\epsilon > 0$,

$$n^{-1} \log(\gamma(G, SP)) \leq -\psi_{ext}(\nu) + \epsilon,$$

uniformly in $l \geq \delta n$, $n \geq n_0(\rho, \delta, \epsilon)$. When it is clear in the context what C is, we will often omit C in the notations.

Similarly, under the parameter C , Section VII below shows that the decay exponent for the internal angle $\beta(F, G)$ is $\psi_{int}(\nu; \rho, \delta)$, which is defined in Section VII. Since $k \sim \rho\delta n$, $l \sim \nu n$, we will have the scaling

$$n^{-1} \log(\beta(F, G)) = -\psi_{int}(\nu; \rho, \delta) + R_2,$$

where the remainder $R_2 = o(1)$ uniformly in $l \geq \delta n$ when $n \geq n_0(\rho, \delta, \epsilon)$ is a large enough natural number.

In summary, under a given $C > 1$, for any fixed choice of ρ, δ , for $\epsilon > 0$, and for $n \geq n_0(\rho, \delta, \epsilon)$,

$$n^{-1} \log(Ds) \leq \psi_{com}(\nu; \rho, \delta) - \psi_{int}(\nu; \rho, \delta) - \psi_{ext}(\nu; \rho, \delta) + 3\epsilon, \quad (24)$$

holds uniformly over the sum parameter s in (14).

In the rest of this paper, when the parameters ρ, δ and C are clear from the context, we will omit them from the notations for the combinatorial, internal and external exponents.

A. Characterizing $\rho_N(\delta, C)$

Continuing to follow [Don06b], we define the *net exponent* $\psi_{net} = \psi_{com}(\nu; \rho, \delta) - \psi_{int}(\nu; \rho, \delta) - \psi_{ext}(\nu; \rho, \delta)$. We will know that the components of ψ_{net} are all continuous over sets $\rho \in [\rho_0, 1]$, $\delta \in [\delta_0, 1]$, $\nu \in [\delta, 1]$, and ψ_{net} is also continuous over these regions.

Definition 1: Let $\delta \in (0, 1]$. The critical proportion $\rho_N(\delta, C)$ is the supremum of $\rho \in [0, 1]$ satisfying

$$\psi_{net}(\nu; \rho, \delta) < 0, \quad \nu \in [\delta, 1].$$

Continuity of ψ_{net} shows that if $\rho < \rho_N$ then, for some $\epsilon > 0$,

$$\psi_{net}(\nu; \rho, \delta) < -\epsilon, \quad \nu \in [\delta, 1].$$

Combine this with (24), for all $s = 0, 2, \dots, (n - m)/2$ and all $n > n_0(\delta, \rho, \epsilon)$,

$$n^{-1} \log(D_s) \leq -\epsilon.$$

Note that if this negative exponent condition holds, we will have the results in Theorem 2.

In the next section, we will specify the exponents ψ_{int} and ψ_{ext} for the internal angles and external angles respectively, and we will discuss properties of $\rho_N(\delta, C)$.

V. CHARACTERIZATIONS OF ANGLE EXPONENTS

A. Exponent for External Angle

Let G denote the cumulative distribution function of a half-normal $HN(0, 1/2)$ random variable, i.e. a random variable $X = |Z|$ where $Z \sim N(0, 1/2)$, and $G(x) = \text{Prob}\{X \leq x\}$, where the density function $g(x) = 2/\sqrt{\pi} \exp(-x^2)$ and thus $G(x)$ is the error function

$$G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy. \quad (25)$$

For $\nu \in (0, 1]$, define x_ν as the solution of

$$\frac{2xG(x)}{g(x)} = \frac{1 - \nu}{\nu'}, \quad (26)$$

where

$$\nu' = (C^2 - 1)\rho\delta + \nu.$$

Because $xG(x)$ is a smooth strictly increasing function, which goes to 0 as $x \rightarrow 0$ and behaves close to x as $x \rightarrow \infty$, and because $g(x)$ is strictly decreasing, the function $2xG(x)/g(x)$ is a strictly increasing function. So x_ν is a well-defined, smooth, and decreasing function of ν .

We have $x_\nu \rightarrow 0$ as $\nu \rightarrow 1$ and $x_\nu \sim \sqrt{\log((1-\nu)/\nu')}$ as $\nu \rightarrow 0$. Define now

$$\psi_{ext}(\nu) = -(1-\nu) \log(G(x_\nu)) + \nu x_\nu^2.$$

This function is smooth on the interior of $(0, 1)$, with endpoints $\psi_{ext}(1) = 0$, $\psi_{ext}(0) = 0$. When $C = 1$, we have the asymptotic [Don06b]

$$\psi_{ext}(\nu) \sim \nu \log\left(\frac{1}{\nu}\right) - \frac{1}{2}\nu \log(\log(\frac{1}{\nu})) + o(\nu), \nu \rightarrow 0. \quad (27)$$

B. Exponent for Internal Angle

Closely following [Don06b], take Y as a standard half-normal random variable $HN(0, 1)$. From standard calculations, we know that its cumulant generating function $\Lambda(s) = \log(E(\exp(sY)))$ is given by

$$\Lambda(s) = \frac{s^2}{2} + \log(2\Phi(s)),$$

where Φ is the usual cumulative distribution function of a standard Normal $N(0, 1)$. So the large deviation rate function of the cumulant generating function Λ^* is defined as

$$\Lambda^*(y) = \max_s sy - \Lambda(s).$$

From the large deviation theory, this function is smooth and convex on $(0, \infty)$, strictly positive except being equal to 0 at $\mu = E(Y) = \sqrt{2/\pi}$. For $\gamma' \in (0, 1)$ let

$$\xi_{\gamma'}(y) = \frac{1-\gamma'}{\gamma'} y^2/2 + \Lambda^*(y), \quad (28)$$

where we define

$$\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}.$$

The function $\xi_{\gamma'}(y)$ is strictly convex and positive on $(0, \infty)$ and has a unique minimum at a unique $y_{\gamma'}$ in the interval $(0, \sqrt{2/\pi})$. Then we have the internal angle exponent as

$$\psi_{int}(\nu; \rho, \delta) = \xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta) + \log(2)(\nu - \rho\delta). \quad (29)$$

For fixed ρ, δ , Λ_{int} is continuous in $\nu \geq \delta$. Most importantly, in the section below, we get the asymptotic formula

$$\xi_{\gamma'}(y_{\gamma'}) \sim \frac{1}{2} \log\left(\frac{1-\gamma'}{\gamma'}\right), \gamma' \rightarrow 0 \quad (30)$$

Because $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}$, (30) means for small $\rho, \nu \in [\delta, 1]$ and any given $\eta > 0$

$$\psi_{int}(\nu, \rho\delta) \geq \left(\frac{1}{2} \cdot \log\left(\frac{1-\gamma'}{\gamma'}\right)(1-\eta) + \log(2)\right)(\nu - \rho\delta). \quad (31)$$

C. Properties of $\rho_N(\delta, C)$

We now consider the combined behavior of ψ_{com} , ψ_{int} and ψ_{net} . We think of these as functions of ν with ρ , δ as parameters. ψ_{com} is the exponent of a growing function which must be outweighed by the sum of the other two exponents: $\psi_{int} + \psi_{net}$.

The asymptotic relations (27) and (30) allow us to see the following key facts about $\rho_N(\delta, C)$, the proofs of which are given in the appendix.

Lemma 3: For any $\delta > 0$ and any $C > 1$, we have

$$\rho_N(\delta, C) > 0, \delta \in (0, 1). \quad (32)$$

Generalizing the result in [Don06b] for $\rho_N(\delta, 1)$, one can show the asymptotic of $\rho_N(\delta, C) \rightarrow 0$ as $\delta \rightarrow 0$.

Lemma 4: For all $\eta > 0$ and any $C > 1$,

$$\rho_N(\delta, C) \geq \log\left(\frac{1}{\delta}\right)^{-(1+\eta)}, \quad \delta \rightarrow 0. \quad (33)$$

Finally, we have the lower and upper bounds for $\rho_N(\delta, C)$, which shows the scaling bounds for $\rho_N(\delta, C)$ as a function of C .

Lemma 5: When $C \geq 1$, for any fixed $\delta > 0$,

$$\Omega\left(\frac{1}{C^2}\right) \leq \rho_N(\delta, C) \leq \frac{1}{C+1}, \quad (34)$$

where $\Omega(\frac{1}{C^2}) \leq \rho_N(\delta, C)$ means that there exists a constant $\iota(\delta)$,

$$\frac{\iota(\delta)}{C^2} \leq \rho_N(\delta, C), \quad \text{as } C \rightarrow \infty,$$

where we can take $\iota(\delta) = \rho_N(\delta, 1)$.

VI. DERIVING THE EXTERNAL ANGLE EXPONENTS

In the previous section, we described how to compute the external and internal angle exponents, and we will give the derivations which justify the computations of the two exponents. First, we start justifying the computation of ψ_{ext} given in Section V.

Lemma 6: Fix $\delta, \epsilon > 0$

$$n^{-1} \log(\gamma(G, \mathbf{SP})) < -\psi_{net}(l/n) + \epsilon_1, \quad (35)$$

uniformly in $l \geq \delta n$, when n is large enough.

Proof: In the appendix, we derived the explicit integral formula for the external angle:

$$\gamma(G, \text{SP}) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx. \quad (36)$$

After a changing of integral variables, we have

$$\gamma(G, \text{SP}) = \sqrt{\frac{(C^2-1)k+l}{\pi}} \int_0^\infty e^{-((C^2-1)k+l)x^2} \left(\frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \right)^{n-l} dx. \quad (37)$$

Inside the parenthesis is the error function G from (25). Let $\nu = \frac{l}{n}$, $\nu' = (C^2-1)\rho\delta + \nu$ then the integral formula can be written as

$$\sqrt{\frac{n\nu'}{\pi}} \int_0^\infty e^{-n\nu'x^2 + n(1-\nu)\log(G(x))} dx. \quad (38)$$

To look at the asymptotic behavior of (38), following the same methodology as in [Don06b], we use Laplace's method. We define

$$f_{\rho,\delta,\nu,n}(y) = e^{-n\psi_{\rho,\delta,\nu}(y)} \cdot \sqrt{\frac{n\nu'}{\pi}} \quad (39)$$

with

$$\psi_{\rho,\delta,\nu}(y) = \nu'y^2 - (1-\nu)\log(G(y))$$

We will develop expressions for the second and third derivatives of the function $\psi_{\rho,\delta,\nu}$. Applying Laplace's method to $\psi_{\rho,\delta,\nu}$ gives the following lemma, where we will defer the proof to later parts of the paper.

Lemma 7: For $\nu \in (0, 1)$, let x_ν denote the minimizer of $\psi_{\rho,\delta,\nu}$. Then

$$\int_0^\infty f_{\rho,\delta,\nu,n}(x) dx \leq e^{-n\psi_{\rho,\delta,\nu}(x_\nu)(1+R_n(\nu))},$$

where for $\delta, \eta > 0$,

$$\sup_{\nu \in [\delta, 1-\eta]} R_n(\nu) = o(1) \text{ as } n \rightarrow \infty,$$

and x_ν is exactly the same x_ν defined earlier in (26).

Recall that the defined exponent ψ_{ext} is given by

$$\psi_{ext}(\nu; \rho, \delta) = \psi_{\rho,\delta,\nu}(x_\nu). \quad (40)$$

Using the definition of $\psi_{\rho,\delta,\nu}(x_\nu)$ and (40), it is not hard to see, as $\nu \rightarrow 1$, $x_\nu \rightarrow 0$ and $\psi_{ext}(\nu) \rightarrow 0$. For any given $\epsilon_1 > 0$ in Lemma 6, there is a largest $\nu_{\epsilon_1} < 1$ with $\psi_{ext}(\nu_{\epsilon_1}) = \epsilon_1$. Note that $\gamma(G, \text{SP}) \leq 1$, so for $l > \nu_{\epsilon_1}n$,

$$n^{-1} \log(\gamma(G, \text{SP})) \leq 0 < -\psi_{ext}(\nu) + \epsilon_1,$$

for $n \geq 1$. Consider now $l \in [\delta n, \nu_{\epsilon_1}n]$, based on (38),

$$\gamma(G, \text{SP}) = \int_0^\infty f_{\rho,\delta,\nu,n}(y) dx.$$

From Lemma 7, as $n \rightarrow \infty$, uniformly for $l \in [\delta n, \nu_{\epsilon_1}n]$,

$$n^{-1} \log(\gamma(G, \text{SP})) = \psi_\nu(x_\nu) + o(1),$$

where we have abbreviated $\psi_{\rho,\delta,\nu}(\cdot)$ to $\psi_\nu(\cdot)$ for fixed ρ and δ .

So from the identity (40), we get

$$n^{-1} \log(\gamma(G, \text{SP})) \leq -\psi_{net}(l/n) + o(1). \quad (41)$$

Then Lemma 6 follows. ■

Now it remains to prove the uniformity result for Laplace's method in Lemma 7. We will follow the same line of reasoning given in [Don06b]. First, we state explicitly the key lemma from [Don06b].

Lemma 8: [Don06b] Let $\psi(x)$ be convex in x and belong to the differentiability class \mathbb{C}^2 (the second derivative exists and is continuous) on an interval I and suppose that it takes its minimum at an interior point $x_0 \in I$, where $\psi''(x_0) > 0$ and that in a vicinity $(x_0 - \epsilon, x_0 + \epsilon)$ of x_0 :

$$|\psi''(x) - \psi''(x_0)| \leq D|\psi''(x_0)||x - x_0|. \quad (42)$$

Let $\bar{\psi}$ be the quadratic approximation $\psi(x_0) + \psi''(x_0)(x - x_0)^2/2$. Then

$$\int_I \exp(-n\psi(x)) dx \leq \int_{-\infty}^{\infty} \exp(-n\bar{\psi}(x)) dx \cdot (S_{1,n} + S_{2,n})$$

where

$$S_{1,n} = \exp(n\psi''(x_0)D\epsilon^3/6)$$

$$S_{2,n} = 2/\left(n\epsilon(2\pi|\psi''(0)|)^{\frac{1}{2}}(1 - \frac{1}{2}D\epsilon^2)\right)$$

The constant D in this lemma can be a scaled third derivative, since if ψ is \mathbb{C}^3 , we can take

$$D = \sup_{(x_0-\epsilon, x_0+\epsilon)} \psi^{(3)}(x)/\psi''(x).$$

Based on Lemma 8, we can derive the uniformity in Lemma 7. In fact, if we pick $\epsilon_n = n^{-\frac{2}{5}}$ and let $n \geq n_1(\psi''(x_0), D)$, where $n_1(\psi''(x_0), D)$ is a number depending only on $\psi''(x_0)$ and D , we can use

$$\int_I e^{-n\psi(x)} dx \leq \int_{-\infty}^{\infty} e^{-n\bar{\psi}(x)} dx \cdot (1 + o(1)) \quad (43)$$

Here the term $o(1)$ is uniform over any collection of convex functions with a given $\psi''(x_0)$ and D . From here to the end of this section, we will abbreviate $\psi_{\rho, \delta, \nu}$ as ψ_ν for the fixed parameter ρ and δ .

Now we consider the collection of convex functions ψ_ν ($\nu \in [\delta, 1 - \eta]$) in Lemma 7. Following the derivations in [Don06b], if we can show that there exists a certain $\epsilon > 0$ so that $\psi''(x_0)$ and D is bounded for the function $\psi_\nu(x)$ uniformly over the range $\nu \in [\delta, 1 - \eta]$, then Lemma 7 holds. Indeed, this is true based on Lemma 9 as given below.

Lemma 9: The function $\psi_\nu(\cdot)$ is smooth with its second derivative at x_ν

$$\psi''_\nu(x_\nu) = 2\nu' + 4x_\nu^2\nu' + \frac{4x_\nu^2\nu'^2}{1 - \nu} \quad (44)$$

and its third derivative at x_ν

$$\psi^{(3)}_\nu(x_\nu) = (1 - \nu) \left((2 - 4x_\nu^2)z - 6x_\nu z^2 - 2z^3 \right) \quad (45)$$

where $z = z_\nu = 2\nu'x_\nu/(1 - \nu)$. We have

$$0 < 2\delta \leq \inf_{\nu \in [\delta, 1]} \psi''_\nu(x_\nu),$$

and

$$\sup_{\nu \in [\delta, 1-\eta]} \psi''_\nu(x_\nu) < \infty.$$

Moreover, for small enough $\epsilon > 0$, the ratio

$$D(\epsilon; \delta, \eta) = \sup_{\nu \in [\delta, 1-\eta]} \sup_{|x-x_\nu| < \epsilon} |\psi^{(3)}_\nu(x)/\psi''_\nu(x)|$$

is finite.

Proof: We can get the following first, second, third derivatives of the function $\psi_\nu(x)$:

$$\psi'_\nu(x) = -(1 - \nu)g/G + 2\nu'x;$$

$$\psi''_\nu(x) = -(1 - \nu)(g'/G - g^2/G^2) + 2\nu';$$

$$\psi^{(3)}_\nu(x) = -(1 - \nu)(g''/G - 3g'g/G^2 + 2g^3/G^3);$$

Because $g' = (-2x)g$, $g'' = (-2 + 4x^2)g$, and

$$g(x_\nu)/G(x_\nu) = \frac{2\nu'x_\nu}{1 - \nu} = z_\nu$$

at the point x_ν , we can immediately have (44) and (45).

Notice that $\psi''_\nu(x_\nu) \geq 2\nu'$, so it is bounded away from zero on any interval $\nu \in [\delta, 1]$, $\delta > 0$. Also, since x_ν is a continuous function bounded away from zero over ν on the interval $[\delta, 1 - \eta]$ ($\delta, \eta > 0$), we have $\psi''_\nu(x_\nu)$ is also bounded above over $[\delta, 1 - \eta]$.

Now as for $\psi^{(3)}$, we note that clearly x_ν and z_ν are continuous functions on $[\delta, 1]$. And both are bounded on the interval $\nu \in [\delta, 1 - \eta]$. As a polynomial in ν , x_ν and z_ν , $\psi^{(3)}_\nu$ is also bounded. If we consider the interval $(x_\nu - \epsilon, x_\nu + \epsilon)$, the boundness of the ratio $D(\epsilon; \delta, \eta)$ also holds uniformly over $\nu \in [\delta, 1 - \eta]$ by inspection if $\epsilon > 0$ is small enough. ■

VII. BOUNDS ON THE INTERNAL ANGLE

In this section, we will show how to get the internal angle decay exponent; namely we will prove the following lemma:

Lemma 10: For $\epsilon > 0$ and $n > n_0(\rho, \delta, \epsilon)$

$$n^{-1} \log(\beta(F, G)) \leq \psi_{int}(l/n; k/l, \delta) + \epsilon,$$

uniformly in $l \geq \delta n$, $k \leq \rho \delta n$, $(l - k) \geq (\nu - \delta \rho)n$.

Using the formula for the internal angle derived in the appendix, we know that

$$-n^{-1} \log(\beta(F, G)) = -n^{-1} \log(B(\frac{1}{1 + C^2 k}, l - k)), \quad (46)$$

where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m' - 1)\alpha' + 1} \pi^{-m'/2} \alpha'^{-1/2} J(m', \theta), \quad (47)$$

with $\theta = (1 - \alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\theta v^2 + 2iv\lambda} dv \right)^{m'} e^{-\lambda^2} d\lambda. \quad (48)$$

To evaluate (46), we need to evaluate the complex integral in $J(m', \theta')$. A saddle point method based on contour integration was sketched for similar integral expressions in [VS92]. A probabilistic method using large deviation theory for evaluating similar integrals was developed in [Don06b]. Both of these two methods can be applied in our case and of course they will produce the same final results. In this paper we will follow the probabilistic method from [Don06b] in this paper. The basic idea is to see the integral in $J(m', \theta')$ as the convolution of $(m' + 1)$ probability densities being expressed in the Fourier domain. In [Don06b], it took mechanical manipulations of the characteristic functions of the normal and half-normal distribution to arrive at this probabilistic method. In the appendix of this paper, we will give a way of deriving the internal angle formula which leads naturally to this probabilistic method and clearly explains its physical meaning.

More explicitly, we have the following lemma:

Lemma 11: let $\theta = (1 - \alpha')/\alpha'$, where $\alpha' = \frac{1}{C^{2k+1}}$. Let T be a random variable with the $N(0, \frac{1}{2})$ distribution, and let $W_{m'}$ be a sum of m' i.i.d. half normals $U_i \sim HN(0, \frac{1}{2\theta})$. Let T and $W_{m'}$ be stochastically independent, and let $g_{T+W_{m'}}$ denote the probability density function of the random variable $T + W_{m'}$. Then⁵

$$B(\alpha', m') = \sqrt{\frac{\alpha'(m' - 1) + 1}{1 - \alpha'}} \cdot 2^{-m'} \cdot \sqrt{\pi} \cdot g_{T+W_{m'}}(0). \quad (49)$$

Applying this probabilistic interpretation and large deviation techniques, it is evaluated as in [Don06b] that

$$g_{T+W_{m'}} \leq \frac{2}{\sqrt{\pi}} \cdot \left(\int_0^{\mu_{m'}} v e^{-v^2 - m' \Lambda^*(\frac{\sqrt{2\theta}}{m'} v)} dv + e^{-\mu_{m'}^2} \right), \quad (50)$$

where Λ^* is the rate function for the standard half-normal random variable $HN(0, 1)$ and $\mu_{m'}$ is the expectation of $W_{m'}$. In fact, the second term in the sum is argued to be negligible [Don06b].

And after taking $y = \frac{\sqrt{2\theta}}{m'} v$, we have an upper bound for the first term:

$$\frac{2}{\sqrt{\pi}} \cdot \frac{m'^2}{2\theta} \cdot \int_0^{\sqrt{2/\pi}} y e^{-m'(\frac{m'}{2\theta})y^2 - m' \Lambda^*(y)} dy. \quad (51)$$

⁵In [Don06b], the term $2^{-m'}$ was $2^{1-m'}$, but we believe that $2^{-m'}$ is the right term.

A. Laplace's Method for ψ_{int}

As we know, m' in the exponent of (51) is defined as $(l - k)$. Similar to evaluating the external angle decay exponent, again we will use Laplace's method in evaluating the internal angle decay exponent. In fact, the function $\xi_{\gamma'}$ of (28) appears in the exponent of (51), with $\gamma' = \frac{\theta}{m' + \theta}$. Since $\theta = \frac{1 - \alpha'}{\alpha'} = C^2 k$, we have

$$\gamma' = \frac{\theta}{m' + \theta} = \frac{C^2 k}{(C^2 - 1)k + l}.$$

Since $k \sim \rho \delta n$, $l \sim \nu n$,

$$\gamma' = \frac{k}{\frac{l}{C^2} + \frac{C^2 - 1}{C^2} k} = \frac{\rho \delta}{\frac{C^2 - 1}{C^2} \rho \delta + \frac{\nu}{C^2}}.$$

Define the function

$$f_{\gamma', m'}(y) = y e^{-m' \xi_{\gamma'}(y)},$$

where $\xi_{\gamma'}(y)$ is the function as defined in (28).

If we apply similar arguments as in proving Lemma 7, we will get the following lemma.

Lemma 12: For $\gamma' \in (0, 1]$ let $y_{\gamma'} \in (0, 1)$ denote the minimizer of $\xi_{\gamma'}$. Then

$$\int_0^\infty f_{\gamma', m'}(x) dx \leq e^{-m' \xi_{\gamma'}(y_{\gamma'})} \cdot R_{m'}(\gamma')$$

where, for $\eta > 0$

$$m'^{-1} \sup_{\gamma' \in [\eta, 1]} \log(R_{m'}(\gamma')) = o(1) \text{ as } m' \rightarrow \infty.$$

This means that

$$g_{T+W_{m'}}(0) \leq e^{-m' \xi_{\gamma'}(y_{\gamma'})} R_{m'}(\gamma').$$

So applying (49), we get

$$n^{-1} \log(\beta(F, G)) \leq (-\xi_{\gamma'}(y_{\gamma'}) - \log(2)) (\nu - \rho \delta) + o(1),$$

where the $o(1)$ is uniform over the range of k and l .

B. Asymptotics of $\xi_{\gamma'}$

As in our previous discussion, we define $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}$, so γ' can take any value in the range $(0, 1]$. Now we are interested in studying the asymptotics of $\xi_{\gamma'}(y_{\gamma'})$ as $\gamma' \rightarrow 0$. As in [Don06b], using the convex duality associated to the cumulant generating function $\Lambda(s)$ and its dual Λ^* , we have

$$y = \Lambda'(s), \quad s = (\Lambda^*)'(y),$$

defining a one-one relationship $s = s(y)$ and $y = y(s)$ between $s < 0$ and $0 < y < \sqrt{\frac{2}{\pi}}$.

From these relations, following the same line of reasoning in [Don06b], we can get the minimizer $y_{\gamma'}$ of $\xi_{\gamma'}$

$$\frac{1 - \gamma'}{\gamma'} y_{\gamma'} = -s_{\gamma'}, \quad (52)$$

where $s_{\gamma'} = s(y_{\gamma'})$.

Because the cumulant generating function for a standard half-normal $HN(0, 1)$ random variable Y is $\Lambda(s) = s^2/2 + \log(2\Phi(s))$, where ϕ and Φ are the standard density and cumulative distributions, we have from $y = \Lambda'(s)$ that

$$y(s) = s \cdot \left(1 - \frac{1}{M(s)}\right), \quad s < 0 \quad (53)$$

where the function of $M(s)$ is defined on $s < 0$ with $0 < M(s) < 1$ and $M(s) \rightarrow 1$ as $s \rightarrow -\infty$ so that

$$\Phi(s) = M(s) \cdot \frac{\phi(s)}{|s|}.$$

Combining (52) and (53), we know that

$$M(s_{\gamma'}) = 1 - \gamma'. \quad (54)$$

Further, we can derive that

$$\xi_{\gamma'}(y_{\gamma'}) = -\frac{1}{2} y_{\gamma'}^2 \frac{1 - \gamma'}{\gamma'} - \log(2/\pi)/2 + \log(y_{\gamma'}/\gamma'). \quad (55)$$

So by the property of the function $M(s)$ and (54), as $\gamma' \rightarrow 0$, $s_{\gamma'} \rightarrow -\infty$, we have

$$Ee^{sY} = \frac{2}{2\sqrt{\pi}} \frac{M(s)}{|s|} \sim \frac{2}{2\sqrt{\pi}} \frac{1}{|s|}, \quad s \rightarrow -\infty.$$

By taking the logarithm for $\Lambda(s)$, $\Lambda(s) \sim -\log|s|$ and $\Lambda'(s) \sim -\frac{1}{s}$ as $s \rightarrow -\infty$. So by $y = \Lambda'(s)$, we have

$$y(s) \sim \frac{1}{|s|} \quad s \rightarrow -\infty,$$

and by combining this with (52), as $\gamma' \rightarrow 0$,

$$y_{\gamma'} \sim \sqrt{\frac{\gamma'}{1 - \gamma'}}.$$

VIII. “WEAK”, “SECTIONAL” AND “STRONG” ROBUSTNESS

So far, we have discussed the robustness of ℓ_1 minimization for sparse signal recovery in the “strong” case, namely we required robust signal recovery for all the approximately k -sparse signal vectors \mathbf{x} . But in applications or performance analysis, we are also often interested in the signal recovery robustness in weaker senses. As we shall see, the framework given in the previous sections can be naturally extended to the analysis of other notions of robustness for sparse signal recovery, resulting in a coherent analysis scheme. For example, we hope to get a tighter performance bound for a particular signal vector instead of a more general, but looser, performance bound for all the possible signal vectors. In this section, we will present our null space conditions on the matrix A to guarantee the performance of the programming (2) in the “weak”, “sectional” and “strong” senses. Here the robustness in the “strong” sense is exactly the robustness we discussed in the previous sections.

Theorem 3: Let A be a general $m \times n$ measurement matrix, \mathbf{x} be an n -element vector and $\mathbf{y} = A\mathbf{x}$. Denote K as a subset of $\{1, 2, \dots, n\}$ such that its cardinality $|K| = k$ and further denote $\overline{K} = \{1, 2, \dots, n\} \setminus K$. Let \mathbf{w} denote an $n \times 1$ vector. Let $C > 1$ be a fixed number.

- (Weak Robustness) Given a specific set K and suppose that the part of \mathbf{x} on K , namely \mathbf{x}_K is fixed. $\forall \mathbf{x}_{\overline{K}}$, any solution $\hat{\mathbf{x}}$ produced by (2) satisfies

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1} \|\mathbf{x}_{\overline{K}}\|_1$$

and

$$\|(\mathbf{x} - \hat{\mathbf{x}})_{\overline{K}}\|_1 \leq \frac{2C}{C-1} \|\mathbf{x}_{\overline{K}}\|_1,$$

if and only if $\forall \mathbf{w} \in \mathbb{R}^n$ such that $A\mathbf{w} = 0$, we have

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1 \geq \|\mathbf{x}_K\|_1; \quad (56)$$

- (Sectional Robustness) Given a specific set $K \subseteq \{1, 2, \dots, n\}$. Then $\forall \mathbf{x} \in \mathbb{R}^n$, any solution $\hat{\mathbf{x}}$ produced by (2) will satisfy

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1} \|\mathbf{x}_{\overline{K}}\|_1,$$

if and only if $\forall \mathbf{x}' \in \mathbb{R}^n, \forall \mathbf{w} \in \mathbb{R}^n$ such that $A\mathbf{w} = 0$,

$$\|\mathbf{x}'_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1 \geq \|\mathbf{x}'_K\|_1; \quad (57)$$

- (Strong Robustness) If for all possible $K \subseteq \{1, 2, \dots, n\}$, and for all $\mathbf{x} \in \mathbb{R}^n$, any solution $\hat{\mathbf{x}}$ produced by (2) satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1} \|\mathbf{x}_{\overline{K}}\|_1,$$

if and only if $\forall K \subseteq \{1, 2, \dots, n\}, \forall \mathbf{x}' \in \mathbb{R}^n, \forall \mathbf{w} \in \mathbb{R}^n$ such that $A\mathbf{w} = 0$,

$$\|\mathbf{x}'_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1 \geq \|\mathbf{x}'_K\|_1. \quad (58)$$

Proof: We will first show the sufficiency of the null space conditions for the various definitions of robustness. Let us begin with the “weak” robustness part. Let $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$ and we must have $A\mathbf{w} = A(\hat{\mathbf{x}} - \mathbf{x}) = 0$. From the triangular inequality for ℓ_1 norm and the fact that $\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{w}\|_1$, we have

$$\begin{aligned} & \|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1 \\ & \geq \|\mathbf{w}_{\overline{K}} + \mathbf{x}_{\overline{K}}\|_1 - \|\mathbf{x}_{\overline{K}}\|_1 \\ & \geq \|\mathbf{w}_{\overline{K}}\|_1 - 2\|\mathbf{x}_{\overline{K}}\|_1. \end{aligned}$$

But the condition (56) guarantees that

$$\|\mathbf{w}_{\overline{K}}\|_1 \geq C(\|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1),$$

so we have

$$\|\mathbf{w}_{\overline{K}}\|_1 \leq \frac{2C}{C-1} \|\mathbf{x}_{\overline{K}}\|_1,$$

and

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1} \|\mathbf{x}_{\overline{K}}\|_1.$$

For the “sectional” robustness, again, we let $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$. Then there must exist an $\mathbf{x}' \in \mathbb{R}^n$ such that

$$\|\mathbf{x}'_K + \mathbf{w}_K\|_1 = \|\mathbf{x}'_K\|_1 - \|\mathbf{w}_K\|_1.$$

Following the condition (57), we have

$$\|\mathbf{w}_K\|_1 \leq \left\| \frac{\mathbf{w}_{\overline{K}}}{C} \right\|_1.$$

Since

$$\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{w}\|_1,$$

following the proof of Theorem 1, we have

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1} \|\mathbf{x}_{\overline{K}}\|_1.$$

The sufficiency of the condition (58) for strong robustness also follows.

Necessity: Since in the proof of the sufficiency, equalities can be achieved in the triangular equalities, the conditions (56), (57) and (58) are also necessary conditions for the the respective robustness to hold for every \mathbf{x} (otherwise, for certain \mathbf{x} 's, there will be $\mathbf{x}' = \mathbf{x} + \mathbf{w}$ with $\|\mathbf{x}'\|_1 < \|\mathbf{x}\|_1$ while violating the respective robustness definitions. Also, such \mathbf{x}' can be the solution to (2)). The detailed arguments will similarly follow the proof of the second part of Theorem 2. ■

The conditions for “weak”, “sectional” and “strong” robustness seem to be very similar, and yet there are indeed huge differences. The “weak” robustness condition is for \mathbf{x} with a specific \mathbf{x}_K on a specific subset K , the “sectional” robustness condition is for \mathbf{x} with all possible \mathbf{x}_K 's on a specific subset K , and the “strong” robustness conditions are for \mathbf{x} 's with all possible \mathbf{x}_K 's on all possible subsets. Basically, the “weak” robustness condition (56) guarantees that the ℓ_1 norm of $\hat{\mathbf{x}}_K$ is not too far away from the ℓ_1 norm of \mathbf{x}_K and the error vector $\mathbf{w}_{\overline{K}}$ is small in ℓ_1 norm when $\|\mathbf{x}_{\overline{K}}\|_1$ is small. Notice that if we define

$$\kappa = \max_{A\mathbf{w}=0, \mathbf{w} \neq 0} \frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}_{\overline{K}}\|_1},$$

then

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2C(1+\kappa)}{C-1} \|\mathbf{x}_{\overline{K}}\|_1.$$

That means, if κ is not ∞ for a measurement matrix A , $\|\mathbf{x} - \hat{\mathbf{x}}\|_1$ is also small when $\|\mathbf{x}_{\overline{K}}\|_1$ is small. Indeed, it is not hard to see that, for a given matrix A , $\kappa < \infty$ as long as the rank of matrix A_K is equal to $|K| = k$, which is generally satisfied for $k < m$.

While the “weak” robustness condition is only for one specific signal \mathbf{x} , the “sectional” robustness condition instead guarantees that given *any* approximately k -sparse signal mainly supported on the subset K , the ℓ_1 -minimization gives a solution $\hat{\mathbf{x}}$ close to the original signal by satisfying (3). When we measure an approximately k -sparse signal \mathbf{x} (the support of the k largest-magnitude components is fixed though unknown to the decoder) using a randomly generated measurement matrix A , the “sectional” robustness conditions characterize the probability that the ℓ_1 minimization solution satisfies (3) for *any* signals for the set K . If that probability goes to 1 as $n \rightarrow \infty$ for any subset K , we know that there exist measurement matrices A ’s that guarantee (3) on “almost all” support sets (namely, (3) is “almost always” satisfied). The “strong” robustness condition instead guarantees the recovery for approximately sparse signals mainly supported on *any* subset K . The “strong” robustness condition is useful in guaranteeing the decoding bound *simultaneously* for *all* approximately k -sparse signals under a single measurement matrix A .

Interestingly, after we take $C = 1$ and let (56), (57) and (58) take strict inequality for all $\mathbf{w} \neq 0$ in the null space of A , the conditions (56), (57) and (58) are also sufficient and necessary conditions for unique exact recovery of ideally k -sparse signals in “weak”, “sectional” and “strong” senses [Don06b], namely the unique exact recovery of a specific ideally k -sparse signal, the unique exact recoveries of all ideally k -sparse signal on a specific support set K and the unique exact recoveries of all ideally k -sparse signal on all possible support sets K . In fact, if $\|\mathbf{x}_{\bar{K}}\|_1 = 0$, from similar triangular inequality derivations in Theorem 1, we have $\hat{\mathbf{x}} = \mathbf{x}$ under all the three conditions.

For a given value $\delta = \frac{m}{n}$ and any value $C \geq 1$, we will determine the value of feasible $\zeta = \frac{k}{n}$ for which there exist a sequence of A ’s such that these three conditions are satisfied when $n \rightarrow \infty$ and $\frac{m}{n} = \delta$. As manifested by the statements of the three conditions (56), (57) and (58) and the previous discussions in Section III, we can naturally extend the Grassmann angle approach to analyze the bounds for the probabilities that (56), (57) and (58) fail. Here we will denote these probabilities as P_1 , P_2 and P_3 respectively. Note that there are $\binom{n}{k}$ possible support sets K and there are 2^k possible sign patterns for signal \mathbf{x}_K . From previous discussions, we know that the event that the condition (56) fails is the same for all \mathbf{x}_K ’s of a specific support set and a specific sign pattern. Then following the same line of reasoning as in Section III, we

have

$$P_1 = P_{K,-} \quad (59)$$

$$P_2 \leq 2^k \times P_1, \quad (60)$$

$$P_3 \leq \binom{n}{k} \times 2^k \times P_1, \quad (61)$$

where $P_{K,-}$ is the probability as in (10).

We have the following lemma about the P_1 , P_2 and P_3 :

Lemma 13: For any $C > 1$, we define $\zeta_W(\delta)$, $\zeta_{Sec}(\delta)$, and $\zeta_S(\delta)$ to be the largest fraction $\zeta = \frac{k}{n}$ such that the condition (56) (57) and (58) are satisfied with overwhelming probability as $n \rightarrow 0$ if we sample the $(n - m)$ -dimensional null space uniformly, where $\frac{m}{n} = \delta$. Then

$$\zeta_W(\delta) > 0,$$

$$\zeta_{Sec}(\delta) > 0,$$

$$\zeta_S(\delta) > 0$$

for any $C > 1$ and $\delta > 0$. Also,

$$\lim_{\delta \rightarrow 1} \zeta_W(\delta) = 1$$

for any $C > 1$.

The proof of this lemma is listed in the appendix. It is worthwhile mentioning that the formula for P_1 is exact since there is no union bound involved and so the threshold bound for the “weak” robustness is tight. In a short summary, the results in this section suggest that even if k is very close to the weak threshold for ideally sparse signals, we can still have robustness results for approximately sparse signals while the results using restricted isometry conditions [CRT05] may suggest smaller sparsity level for recovery robustness. This is the first such a kind of result. The numerical results of ζ making sure that P_1 , P_2 , P_3 converge to zero overwhelmingly are presented in Section X.

IX. ANALYSIS OF ℓ_1 MINIMIZATION UNDER NOISY MEASUREMENTS

In the previous sections, we have analyzed the ℓ_1 minimization algorithm for decoding general signals. In this section, we will discuss the effect of noisy measurements on the ℓ_1 minimization of general signals, using the null space characterization.

Theorem 4: Assume that A is a general $m \times n$ measurement matrix A with rank m and its minimum nonzero singular value is denoted as σ_{\min} . Further, assume that $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, with its ℓ_2 -norm $\|\mathbf{b}\| \leq \epsilon$, and that \mathbf{w} is an $n \times 1$ vector. Let K be any subset of $\{1, 2, \dots, n\}$ such that its cardinality $|K| = k$ and let K_i denote the i -th element of K . Further, let $\overline{K} = \{1, 2, \dots, n\} \setminus K$. Then the solution $\hat{\mathbf{x}}$ produced by (2) will satisfy

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1} \|\mathbf{x}_{\overline{K}}\|_1 + \frac{(3C+1)\sqrt{n}\epsilon}{(C-1)\sigma_{\min}}$$

with $C > 1$, if $\forall \mathbf{w} \in \mathbb{R}^n$ such that

$$A\mathbf{w} = 0$$

and for all the subsets K with $|K| = k$, we have

$$C \sum_{i=1}^k |\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k} |\mathbf{w}_{\overline{K}_i}|. \quad (62)$$

Proof: Since

$$\mathbf{y} = A\mathbf{x} + \mathbf{b},$$

we can write

$$\mathbf{y} = A\mathbf{x}^*,$$

where

$$\|\mathbf{x}^* - \mathbf{x}\| \leq \frac{\epsilon}{\sigma_{\min}}.$$

By the Cauchy-Schwarz inequality, we have

$$\|\mathbf{x}^* - \mathbf{x}\|_1 \leq \frac{\sqrt{n}\epsilon}{\sigma_{\min}}.$$

Suppose the matrix A has the claimed null space property. Now the solution $\hat{\mathbf{x}}$ of (2) satisfies $\|\hat{\mathbf{x}}\|_1 \leq \|\mathbf{x}^*\|_1$. Since $A\hat{\mathbf{x}} = \mathbf{y}$, it easily follows that $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}^*$ is in the null space of A . Therefore we can further write $\|\mathbf{x}^*\|_1 \geq \|\mathbf{x}^* + \mathbf{w}\|_1$. Using the triangular inequality for the ℓ_1 norm we obtain

$$\begin{aligned} \|\mathbf{x}_K^*\|_1 + \|\mathbf{x}_{\overline{K}}^*\|_1 &= \|\mathbf{x}^*\|_1 \\ &\geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x}^* + \mathbf{w}\|_1 \\ &\geq \|\mathbf{x}_K^*\|_1 - \|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\overline{K}}\|_1 - \|\mathbf{x}_{\overline{K}}^*\|_1 \\ &\geq \|\mathbf{x}_K^*\|_1 - \|\mathbf{x}_{\overline{K}}^*\|_1 + \frac{C-1}{C+1} \|\mathbf{w}\|_1, \end{aligned}$$

where the last inequality is from the claimed null space property. Relating the first equality and the last inequality above, we have $2\|\mathbf{x}_K^*\|_1 \geq \frac{(C-1)}{C+1}\|\mathbf{w}\|_1$.

Since

$$\|\mathbf{x}_K^*\|_1 \leq \|\mathbf{x}_K\|_1 + \|\mathbf{x}^* - \mathbf{x}\|_1,$$

we get

$$\begin{aligned} \|\mathbf{w}\|_1 &\leq \frac{2(C+1)}{C-1}\|\mathbf{x}_K^*\|_1 \\ &\leq \frac{2(C+1)}{C-1}\|\mathbf{x}_K\|_1 + \frac{2(C+1)}{C-1}\|\mathbf{x}^* - \mathbf{x}\|_1. \end{aligned}$$

From the triangular inequality,

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \|\mathbf{x} - \mathbf{x}^*\|_1 + \|\mathbf{w}\|_1 \tag{63}$$

$$\leq \frac{2(C+1)}{C-1}\|\mathbf{x}_K\|_1 + \frac{3C+1}{C-1}\|\mathbf{x}^* - \mathbf{x}\|_1, \tag{64}$$

$$\leq \frac{2(C+1)}{C-1}\|\mathbf{x}_K\|_1 + \frac{(3C+1)\sqrt{n}\epsilon}{(C-1)\sigma_{\min}}. \tag{65}$$

■

If the elements in the measurement matrix A are i.i.d. as the unit real Gaussian random variables $N(0, 1)$, following upon the work of Marchenko and Pastur [MP67], Geman[Gem80] and Silverstein [Sil85] proved that for $m/n = \delta$, as $n \rightarrow \infty$,

$$\frac{1}{\sqrt{n}}\sigma_{\min} \rightarrow 1 - \sqrt{\delta}$$

almost surely as $n \rightarrow \infty$.

Then almost surely as $n \rightarrow \infty$, $\frac{(3C+1)\sqrt{n}\epsilon}{(C-1)\sigma_{\min}} \rightarrow \frac{(3C+1)\epsilon}{(C-1)(1-\sqrt{\delta})}$. So in this case, we have $\|\mathbf{x}^* - \mathbf{x}\|_1$ is upper-bounded by some constant times ϵ . It is also worth mentioning that the error bound derived above is for a plain ℓ_1 minimization optimization programming, which does not use any prior knowledge of the magnitudes of the noise in the computations, while the error bounds in the literatures often assume that such information is known and is used in the convex programming algorithms. To get an error bound in terms of ℓ_2 norm, we can invoke the almost Euclidean property of the null space, namely every vector \mathbf{w} has an ℓ_2 norm scaling as $O(\frac{1}{\sqrt{n}})$ of its ℓ_1 norm. Though we choose not to do it in detail in this paper, it is easy to see that the error bound here has the same scaling in ϵ as the analysis through the restricted isometry property [CRT06]; however, this analysis is warranted even when the cardinality $|K|$ of the set K is

much larger than the known cardinality bounds for the restricted isometry property. It is also possible to extend the concepts of “weak”, “sectional” and “strong” robustness analysis to noisy measurements, which will also similarly show that even if the cardinality of the set K is very close to the “weak” threshold for the ideally sparse signals, we can still have the robustness of ℓ_1 minimization to noisy measurements.

X. NUMERICAL COMPUTATIONS ON THE BOUNDS OF ζ

In this section, we will numerically evaluate the performance bounds on $\zeta = \frac{k}{n}$ such that the conditions (7), (56), (57) and (58) are satisfied with overwhelming probability as $n \rightarrow \infty$.

First, we know that the condition (7) fails with probability

$$P \leq \binom{n}{k} \times 2^k \times 2 \times \sum_{s \geq 0} \sum_{G \in \mathfrak{S}_{m+1+2s}(\text{SP})} \beta(F, G) \gamma(G, \text{SP}), \quad (66)$$

Recall that we assume $\frac{m}{n} = \delta$, $l = (m + 1 + 2s) + 1$ and $\nu = \frac{l}{n}$. In order to make P overwhelmingly converge to zero as $n \rightarrow \infty$, following the discussions in Section IV, one sufficient condition is to make sure that the exponent for the combinatorial factors

$$\psi_{com} = \lim_{n \rightarrow \infty} \frac{\log \left(\binom{n}{k} 2^k 2^{\binom{n-k}{l-k}} 2^{l-k} \right)}{n} \quad (67)$$

and the negative exponent for the angle factors

$$\psi_{angle} = - \lim_{n \rightarrow \infty} \frac{\log (\beta(F, G) \gamma(G, \text{SP}))}{n} \quad (68)$$

satisfy $\psi_{com} - \psi_{angle} < 0$ uniformly over $\nu \in [\delta, 1)$.

Following [Don06b] we take $m = 0.5555n$. By analyzing the decaying exponents of the external angles and internal angles through the Laplace methods as in Section VI and VII, we can compute the numerical results as shown in Figure 3, Figure 6 and Figure 7. In Figure 3, we show the largest sparsity level $\zeta = \frac{k}{n}$ (as a function of C) which makes the failure probability of the condition (9) approach zero asymptotically as $n \rightarrow \infty$. As we can see, when $C = 1$, we get the same bound $\zeta = 0.095 \times 0.5555 \approx 0.0528$ as obtained for the “weak” threshold for the ideally sparse signals in [Don06b]. As expected, as C grows, the ℓ_1 minimization requires a smaller sparsity level ζ to achieve higher signal recovery accuracy.

In Figure 4, we show the exponents ψ_{com} , ψ_{int} , ψ_{ext} under the parameters $C = 2$, $\delta = 0.5555$ and $\zeta = 0.0265$. For the same set of parameters, in Figure 5, we compare the exponents ψ_{com}

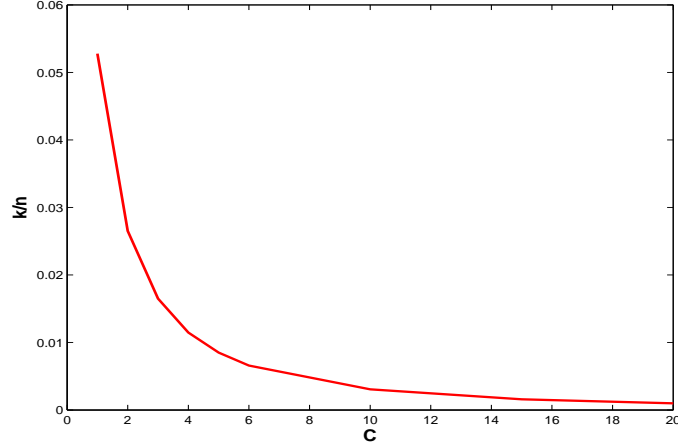


Fig. 3: Allowable sparsity as a function of C (allowable imperfection of the recovered signal is $\frac{2(C+1)\Delta}{C-1}$)

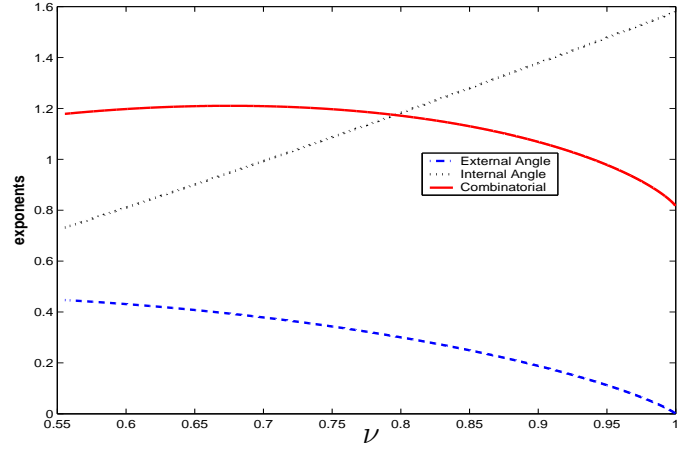


Fig. 4: The Combinatorial, Internal and External Angle Exponents

and ψ_{angle} : the solid curve denotes ψ_{angle} and the dashed curve denotes ψ_{com} . It shows that, under $\zeta = 0.0265$, $\psi_{com} - \psi_{angle} < 0$ uniformly over $\delta \leq \nu \leq 1$. Indeed, $\zeta = 0.0265$ is the bound shown in Figure 3 for $C = 2$. In Figure 6, for the parameter $\delta = 0.5555$, we give the bounds ζ as a function of C for satisfying the signal recovery robustness conditions (56), (57) and (58) respectively in the “weak”, “sectional” and “strong” senses. In Figure 7, fixing $C = 2$, we plot how large $\rho = \zeta/\delta$ can be for different δ ’s while satisfying the signal recovery robustness conditions (56), (57) and (58) respectively in “weak”, “sectional” and “strong” senses.

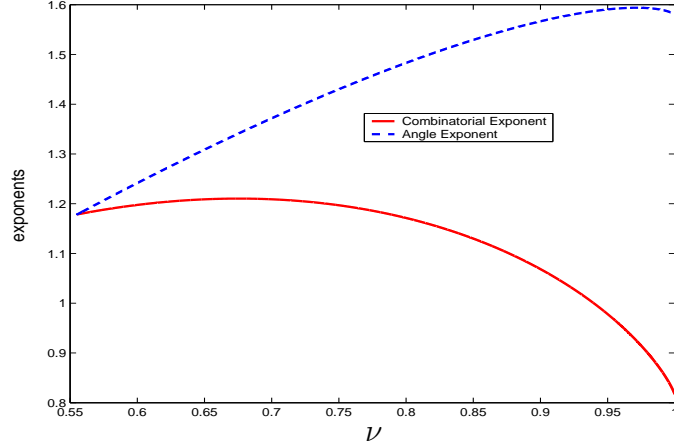


Fig. 5: The Combinatorial Exponents and the Angle Exponents

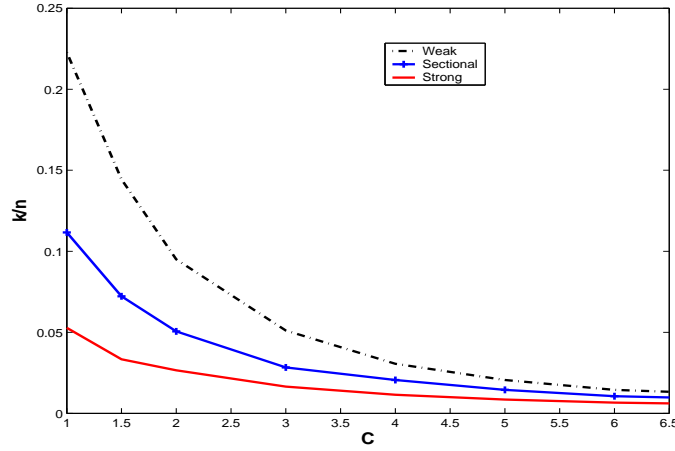


Fig. 6: The Weak, Sectional and Strong Robustness Bounds

XI. CONCLUSION

It is well known that ℓ_1 optimization can be used to recover ideally sparse signals in compressive sensing, if the underlying signal is sparse enough. While for the ideally sparse signals, the results of [Don06b] have given us very sharp bounds on the sparsity threshold the ℓ_1 minimization can recover, sharp bounds for the recovery of general signals or approximately sparse signals were not available.

In this paper we analyzed a null space characterization of the measurement matrices for the performance bounding of ℓ_1 -norm optimization for general signals or approximately sparse.

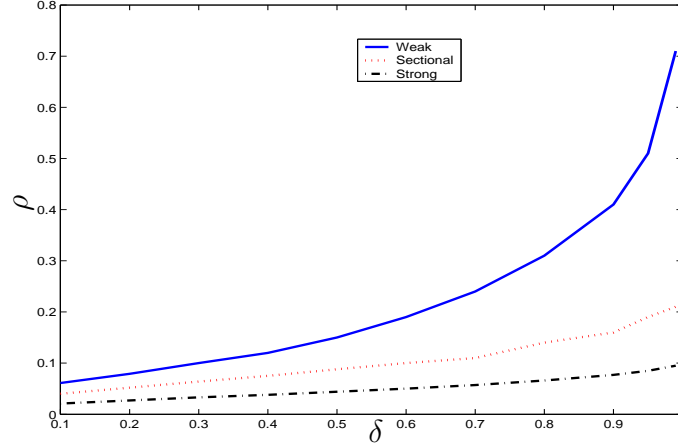


Fig. 7: The Weak, Sectional and Strong Robustness Bounds

Using high-dimensional geometry tools, we give a unified *null space Grassmann angle*-based analytical framework for compressive sensing. This new framework gives sharp quantitative tradeoffs between the signal sparsity parameter and the recovery accuracy of the ℓ_1 optimization for general signals or approximately sparse signals. As expected, the neighborly polytopes result of [Don06b] for ideally sparse signals can be viewed as a special case on this tradeoff curve. It can therefore be of practical use in applications where the underlying signal is not ideally sparse and where we are interested in the quality of the recovered signal. For example, using the results and their extensions in this paper and [Don06b], we are able to give a precise sparsity threshold analysis for weighted ℓ_1 minimization when prior information about the signal vector is available [KXAH09]. In [XKAH10], using the robustness result from this paper, we are able to show that a polynomial-time iterative weighted ℓ_1 minimization algorithm can provably improve over the sparsity threshold of ℓ_1 minimization for interesting classes of signals, even when prior information is not available.

In essence, this work investigates the fundamental “balancedness” property of linear subspaces, and may be of independent mathematical interest. In future work, it is interesting to obtain more accurate analysis for compressive sensing under noisy measurements than presented in the current paper.

XII. APPENDIX

A. Some Concepts in the High Dimensional Geometry

In this part, we will give the explanations of several often used geometric terminologies in this paper for the purpose of quick reference.

1) *the Grassmann Manifold*: The Grassmann manifold $\text{Gr}_i(j)$ refers to the set of i -dimensional subspaces in the j -dimensional Euclidean space \mathbb{R}^j . It is known that there exists a unique invariant measure μ' on $\text{Gr}_i(j)$ such that $\mu'(\text{Gr}_i(j))=1$.

For more facts on the Grassmann manifold, please see [Boo86].

2) *Polytope, Face, Vertex*: A polytope in this paper refers to the convex hull of a finite number points in the Euclidean space. Any extreme point of a polytope is a vertex of this polytope. A face of a polytope is defined as the convex hull of a set of its vertices such that no point in this convex hull is an interior point of the polytope. The dimension of a face refers to the dimension of the affine hull of that face. The book [Grü03] offers a nice reference on the convex polytopes.

3) *Cross-polytope*: The n -dimensional cross-polytope is the polytope of unit ℓ_1 ball, namely it is the set

$$\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_1 = 1\}.$$

The n -dimensional cross-polytope has $2n$ vertices, namely $\pm e_1, \pm e_2, \dots, \pm e_n$, where e_i , $1 \leq i \leq n$, is the unit vector with its i -th coordinate element being 1. Any k extreme points without opposite pairs at the same coordinate will constitute a $(k-1)$ -dimensional face of the cross-polytope. So the cross-polytope will have $2^k \binom{n}{k}$ faces of dimension $(k-1)$.

4) *the Grassmann Angle*: The Grassmann angle for a n -dimensional cone \mathfrak{C} under the Grassmann manifold $\text{Gr}_i(n)$, is the measure of the set of i -dimensional subspaces (over $\text{Gr}_i(n)$) which intersect the cone \mathfrak{C} nontrivially (namely at some other point besides the origin). For more details on the Grassmann angle, internal angle, and external angle, please refer to [Grü68][Grü03][McM75].

5) *the Internal Angle*: An internal angle $\beta(F_1, F_2)$, between two faces F_1 and F_2 of a polytope or a polyhedral cone, is the fraction of the hypersphere S covered by the cone obtained by observing the face F_2 from the face F_1 . The internal angle $\beta(F_1, F_2)$ is defined to be zero when $F_1 \not\subseteq F_2$ and is defined to be one if $F_1 = F_2$. Note the dimension of the hypersphere S here matches the dimension of the corresponding cone discussed. Also, the center of the hypersphere

is the apex of the corresponding cone. All these defaults also apply to the definition of the external angles.

6) *the External Angle*: An external angle $\gamma(F_3, F_4)$, between two faces F_3 and F_4 of a polytope or a polyhedral cone, is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face F_4 at the face F_3 . The external angle $\gamma(F_3, F_4)$ is defined to be zero when $F_3 \not\subseteq F_4$ and is defined to be one if $F_3 = F_4$.

B. Proof of Lemma 1

Proof: The first statement is obvious since multiplying A with a unitary Θ keeps the columns independent and the entries i.i.d. Gaussian.

Now let us look at the proof of the second statement. Consider the Singular Value Decomposition (SVD) $A = U\Sigma V^*$, where U and V have orthonormal columns and Σ is diagonal. Consider now $A\Theta$, for any given deterministic unitary Θ : $A\Theta = U\Sigma V^*\Theta$. This is clearly the SVD of $A\Theta$; in particular, Θ^*V represents the right singular vectors of $A\Theta$. Since A and $A\Theta$ have the same distribution (for all unitary Θ), the same must be true of the right singular vectors V and Θ^*V . Therefore the distribution of V is left-rotationally invariant: $P_V(V) = P_V(\Theta^*V)$. Now the null space of A can be written as $Z = V^\perp X$, where V^\perp is an $n \times (n - m)$ matrix with orthonormal columns that are orthogonal to V , i.e., $V^*V^\perp = 0$, and X is any invertible $(n - m) \times (n - m)$ matrix. Now it is easy to see that if we change V to Θ^*V , for any unitary Θ , we must change V^\perp to Θ^*V^\perp . But since left-multiplication by a unitary Θ^* does not change the distribution of V , left multiplication by a unitary Θ^* must not change the distribution of V^\perp . Thus V^\perp , and by fiat $Z = V^\perp X$, are left-rotationally invariant. Note to simplify the arguments, we have so far assumed that the matrix A is of full rank m , which is true with probability 1. However, we should note that these arguments also work when the matrix A is rank-deficient.

Let G be an $n \times (n - m)$ matrix with i.i.d. $\mathcal{N}(0, 1)$ entries and consider the QR decomposition: $G = QR$, where Q is an $n \times (n - m)$ matrix with orthonormal columns and R is an $(n - m) \times (n - m)$ upper triangular matrix with non-negative diagonals. Then it is well known that Q has a left-rotationally invariant distribution, and that R is a random matrix, independent of Q , whose strictly upper triangular entries are i.i.d. $\mathcal{N}(0, 1)$ and whose i -th diagonal entry is an independent Chi-square random variable with $(n - i + 1)/2$ degrees of freedom [Mui05]. This implies that we can always take the V^\perp obtained from the 2nd statement and post-multiply it by an independent

upper triangular R (with the aforementioned distribution) to obtain a matrix $Z = V^\perp R$ with i.i.d. $\mathcal{N}(0, 1)$ entries. It is always possible to choose a basis Z for the null space such that Z has i.i.d. $\mathcal{N}(0, 1)$ entries. ■

C. Derivation of the Internal Angles

There are two situations in the derivations of the internal angles $\beta(F, G)$ for the skewed cross-polytope: when G is a regular face and when G is the whole skewed cross-polytope SP. These two cases are respectively dealt with in Lemma 14 and Lemma 15.

Lemma 14: Suppose that F is a $(k - 1)$ -dimensional face of the skewed cross-polytope

$$\text{SP} = \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\}$$

supported on the subset K with $|K| = k$. Then the internal angle $\beta(F, G)$ between the $(k - 1)$ -dimensional face F and a $(l - 1)$ -dimensional face G ($F \subseteq G$, $G \neq \text{SP}$) is given by

$$\beta(F, G) = \frac{V_{l-k-1}(\frac{1}{1+C^2k}, l-k-1)}{V_{l-k-1}(S^{l-k-1})}, \quad (69)$$

where $V_i(S^i)$ denotes the i -th dimensional surface measure on the unit sphere S^i , while $V_i(\alpha', i)$ denotes the surface measure for regular spherical simplex with $(i + 1)$ vertices on the unit sphere S^i and with inner product as α' between these $(i + 1)$ vertices. (15) is equal to $B(\frac{1}{1+C^2k}, l-k)$, where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m' - 1)\alpha' + 1} \pi^{-m'/2} \alpha'^{-1/2} J(m', \theta) \quad (70)$$

with $\theta = (1 - \alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\theta v^2 + 2iv\lambda} dv \right)^{m'} e^{-\lambda^2} d\lambda \quad (71)$$

Proof: Without loss of generality, assume that F is a $(k - 1)$ -dimensional face with k vertices as $e_p, 1 \leq p \leq k$, where e_p is the n -dimensional standard unit vector with the p -th element as '1'; and also assume that the $(l - 1)$ -dimensional face G be the convex hull of the l vertices: $e_p, 1 \leq p \leq k$ and $Ce_p, (k + 1) \leq p \leq l$. Then the cone $\text{Con}_{F,G}$ formed by observing the $(l - 1)$ -dimensional face G of the skewed cross-polytope SP from an interior point x^F of the face F is the positive cone of the vectors:

$$Ce_j - e_i, \quad \text{for all } j \in J \setminus K, \quad i \in K, \quad (72)$$

and also the vectors

$$e_{i_1} - e_{i_2}, \quad \text{for all } i_1 \in K, i_2 \in K, \quad (73)$$

where $J = \{1, 2, \dots, l\}$ is the support set for the face G .

So the cone $\text{Con}_{F,G}$ is the direct sum of the linear hull $L_F = \text{lin}\{F - x^F\}$ formed by the vectors in (73) and the cone $\text{Con}_{F^\perp,G} = \text{Con}_{F,G} \cap L_F^\perp$, where L_F^\perp is the orthogonal complement to the linear subspace L_F . Then $\text{Con}_{F^\perp,G}$ has the same spherical volume as $\text{Con}_{F,G}$.

Now let us analyze the structure of $\text{Con}_{F^\perp,G}$. We notice that the vector

$$e_0 = \sum_{p=1}^k e_p$$

is in the linear space L_F^\perp and is also the only such a vector (up to linear scaling) supported on K . Thus a vector \mathbf{x} in the positive cone $\text{Con}_{F^\perp,G}$ must take the form

$$-\sum_{i=1}^k b_i \times e_i + \sum_{i=k+1}^l b_i \times e_i, \quad (74)$$

where $b_i, 1 \leq i \leq l$ are nonnegative real numbers and

$$\begin{aligned} C \sum_{i=1}^k b_i &= \sum_{i=k+1}^l b_i, \\ b_1 &= b_2 = \dots = b_k. \end{aligned}$$

That is to say, the $(l - k)$ -dimensional $\text{Con}_{F^\perp,G}$ is the positive cone of $(l - k)$ vectors a^1, a^2, \dots, a^{l-k} , where

$$a^i = C \times e_{k+i} - \sum_{p=1}^k e_p / k, \quad 1 \leq i \leq (l - k).$$

The normalized inner products between any two of these $(l - k)$ vectors is

$$\frac{\langle a^i, a^j \rangle}{\|a^i\| \|a^j\|} = \frac{k \times \frac{1}{k^2}}{C^2 + k \times \frac{1}{k^2}} = \frac{1}{1 + kC^2}.$$

(In fact, a^i 's are also the vectors obtained by observing the vertices e_{k+1}, \dots, e_l from $Ec = \sum_{p=1}^k e_p / k$, the epicenter of the face F .)

We have so far reduced the computation of the internal angle to evaluating (69), the relative spherical volume of the cone $\text{Con}_{F^\perp,G}$ with respect to the sphere surface S^{l-k-1} . This was

computed as given in this lemma [VS92], [KBH99] for the positive cones of vectors with equal inner products by using a transformation of variables and the well-known formula

$$V_{i-1}(S^{i-1}) = \frac{i\pi^{\frac{i}{2}}}{\Gamma(\frac{i}{2} + 1)},$$

where $\Gamma(\cdot)$ is the usual Gamma function.

Instead, in this paper, we will give a proof of (70) which can directly lead to the probabilistic large deviation method of evaluating the internal angle exponent in [Don06b].

First, we notice that $\text{Con}_{F^\perp, G}$ is a $(l-k)$ -dimensional cone. Also, all the vectors (x_1, \dots, x_n) in the cone $\text{Con}_{F^\perp, G}$ take the form in (74). From [Had79],

$$\begin{aligned} \int_{\text{Con}_{F^\perp, G}} e^{-\|x'\|^2} dx' &= \beta(F, G) V_{l-k-1}(S^{l-k-1}) \\ &\times \int_0^\infty e^{-r^2} r^{l-k-1} dr = \beta(F, G) \cdot \pi^{(l-k)/2}, \end{aligned} \quad (75)$$

where $V_{l-k-1}(S^{l-k-1})$ is the spherical volume of the $(l-k-1)$ -dimensional sphere S^{l-k-1} . Now define $U \subseteq \mathbb{R}^{l-k+1}$ as the set of all nonnegative vectors satisfying:

$$x_p \geq 0, \quad 1 \leq p \leq l-k+1, \quad \sum_{p=2}^{l-k+1} x_p = Ckx_1$$

and define $f(x_1, \dots, x_{l-k+1}) : U \rightarrow \text{Con}_{F^\perp, G}$ to be the linear and bijective map

$$f(x_1, \dots, x_{l-k+1}) = -\sum_{p=1}^k x_1 e_p + \sum_{p=k+1}^l x_{p-k} \times e_p.$$

Then

$$\begin{aligned} &\int_{\text{Con}_{F^\perp, G}} e^{-\|x'\|^2} dx' \\ &= \sqrt{\frac{l + (C^2 - 1)k}{C^2 k}} \int_{\sum_{p=2}^{l-k+1} x_p = Ckx_1, x_p \geq 0, 2 \leq p \leq l-k+1} e^{-\|f(x)\|^2} dx_2 \cdots dx_{l-k+1} \\ &= \sqrt{\frac{l + (C^2 - 1)k}{C^2 k}} \int_{\sum_{p=2}^{l-k+1} x_p = Ckx_1, x_p \geq 0, 2 \leq p \leq l-k+1} e^{-kx_1^2 - x_2^2 - \cdots - x_{l-k+1}^2} dx_2 \cdots dx_{l-k+1} \end{aligned} \quad (76)$$

where $\sqrt{\frac{l + (C^2 - 1)k}{C^2 k}}$ is due to the change of integral variables.

In fact, when

$$x_p \geq 0, \quad 1 \leq p \leq l - k + 1, \quad \sum_{p=2}^{l-k+1} x_p = Ckx_1,$$

the function f is a linear transformation over the variables x_2, \dots, x_{l-k+1} with the following transformation matrix M (disregarding the indices beyond l)

$$\begin{pmatrix} -\frac{1}{Ck} & \cdots & -\frac{1}{Ck} & 1 & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{Ck} & \cdots & -\frac{1}{Ck} & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

It can then be calculated that the Jacobian of this transformation is $\sqrt{\det MM^T} = \sqrt{\frac{l+(C^2-1)k}{C^2k}}$, which accounts for the coefficient appearing in (76).

Now we define a random variable

$$Z = X_2 + X_3 + \cdots + X_{l-k+1} - CkX_1,$$

where X_1, X_2, \dots, X_n are independent random variables, with $X_p \sim HN(0, \frac{1}{2})$, $2 \leq p \leq (n-k+1)$, as half-normal distributed random variables and $X_1 \sim N(0, \frac{1}{2k})$ as a normal distributed random variable. Then by inspection, (76) is equal to

$$\frac{\sqrt{\pi}^{l-k+1}}{2^{l-k}} \times \sqrt{l + (C^2 - 1)k} p_Z(0).$$

where $p_Z(\cdot)$ is the probability density function for the random variable $Z = X_2 + X_3 + \cdots + X_{l-k+1} - CKX_1$ and $p_Z(0)$ is the probability density function $p_Z(z)$ evaluated at the point $z = 0$.

Use the notation

$$G_X(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda x} p_X(x) dx$$

as the characteristic function for any random variable X , where $p_X(x)$ is the probability density function of X . Then from the independence of $X_1, X_2, \dots, X_{l-k+1}$, the characteristic function for Z is equal to

$$G_Z(\lambda) = G_{X_2}(\lambda)^{l-k} \times G_{X_1}(\lambda).$$

Expressing the probability density function $p_Z(x)$ in the Fourier domain, we have

$$p_Z(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_Z(\lambda) d\lambda.$$

Combining this with (75), we can obtain the desired result (70); and, as we already see in the derivations, we naturally arrive at the probabilistic explanation for the internal angle. ■

So far at this point, we have only considered the internal angle $\beta(F, G)$ when G is not the whole skewed cross-polytope. The following lemma discusses the special case when $G = \text{SP}$.

Lemma 15: Suppose F , K and SP are defined in the same way as in the statement of Lemma 14. Then the internal angle $\beta(F, \text{SP})$ between the $(k - 1)$ -dimensional face F and the n -dimensional skewed cross-polytope SP is given by

$$\int_{-\infty}^0 \int_{-\infty}^{\infty} e^{-\frac{C^2 k \lambda^2}{4}} \left(\int_0^{\infty} e^{-\mu^2 + i\mu\lambda} d\mu \right)^{n-k} \frac{2^{n-k-1}}{\sqrt{\pi}^{n-k+2}} e^{-i\lambda z} d\lambda dz.$$

Proof: We use the same set of notations as in the proof of Lemma 14. Without loss of generality, assume $K = \{1, \dots, k\}$, F is the $(k - 1)$ -dimensional face supported on K and $G = \text{SP}$. So the cone $\text{Con}_{F,G}$ is the direct sum of the linear hull $L_F = \text{lin}\{F - x^F\}$ formed by the vectors in (73) and the cone $\text{Con}_{F^\perp, G} = \text{Con}_{F,G} \cap L_F^\perp$, where L_F^\perp is the orthogonal complement to the linear subspace L_F . Then $\text{Con}_{F^\perp, G}$ has the same spherical volume as $\text{Con}_{F,G}$.

Following similar analysis in Lemma 14, the cone $\text{Con}_{F^\perp, G}$ is the positive cone of $2(n - k)$ vectors $a_\pm^1, a_\pm^2, \dots, a_\pm^{n-k}$, where

$$a_\pm^i = \pm C \times e_{k+i} - \sum_{p=1}^k e_p / k, \quad 1 \leq i \leq (n - k).$$

This also means that $\text{Con}_{F^\perp, G}$ is a $(n - k + 1)$ -dimensional cone. Also, all the vectors (b_1, \dots, b_n) in the cone $\text{Con}_{F^\perp, G}$ take the form

$$b_1 = b_2 = \dots = b_k \leq 0, \\ \sum_{p=k+1}^n |b_p| \leq Ck|b_1|.$$

From [Had79],

$$\begin{aligned} \int_{\text{Con}_{F^\perp, G}} e^{-\|x'\|^2} dx' &= \beta(F, G) V_{n-k}(S^{n-k}) \\ &\times \int_0^\infty e^{-r^2} r^{n-k} dr = \beta(F, G) \cdot \pi^{(n-k+1)/2}, \end{aligned} \quad (77)$$

where $V_{n-k}(S^{n-k})$ is the spherical volume of the $(n-k)$ -dimensional sphere S^{n-k} . Now define $U \subseteq \mathbb{R}^{n-k+1}$ as the set of all the vectors taking the form:

$$\{x_1 \geq 0, \sum_{p=2}^{n-k+1} |x_p| \leq Ckx_1\}$$

and define $f(x_1, x_2, \dots, x_{n-k+1}) : U \rightarrow \text{Con}_{F^\perp, G}$ to be the linear and bijective map

$$f(x_1, \dots, x_{n-k+1}) = -\sum_{p=1}^k x_1 e_p + \sum_{p=k+1}^n x_{p-k+1} \times e_p.$$

Then

$$\begin{aligned} \int_{\text{Con}_{F^\perp, G}} e^{-\|x'\|^2} dx' &= \sqrt{k} \int_U e^{-\|f(x)\|^2} dx \\ &= \sqrt{k} \int_0^\infty \int_{\sum_{p=2}^{n-k+1} |x_p| \leq Ckx_1} e^{-kx_1^2 - x_2^2 - \dots - x_{n-k+1}^2} dx_2 \dots dx_{n-k+1} dx_1 \end{aligned} \quad (78)$$

where \sqrt{k} is due to the change of integral variables.

By inspection, (78) is equal to

$$\sqrt{\pi}^{n-k+1} P(X_2 + X_3 + \dots + X_{n-k+1} - CkX_1 \leq 0),$$

where $X_1, X_2, \dots, X_{n-k+1}$ are independent random variables, with $X_p \sim HN(0, \frac{1}{2})$, $2 \leq p \leq (n-k+1)$, as half-normal distributed random variables and $X_1 \sim N(0, \frac{1}{2k})$ as a normal distributed random variable.

Expressing the probability density function of $Z = X_2 + X_3 + \dots + X_{n-k+1} - CKX_1$ in the Fourier domain, we can simplify (78) to

$$\begin{aligned} \int_{-\infty}^0 \int_{-\infty}^\infty e^{-\frac{C^2 k \lambda^2}{4}} \left(\int_0^\infty e^{-\mu^2 + i\mu\lambda} d\mu \right)^{n-k} \\ \frac{2^{n-k-1}}{\sqrt{\pi}} e^{-i\lambda z} d\lambda dz \end{aligned}$$

Combining this with (77) gives us the desired result. ■

D. Derivation of the External Angles

Lemma 16: Suppose that F is a $(k-1)$ -dimensional face of the skewed cross-polytope

$$\text{SP} = \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\}$$

supported on a subset K with $|K| = k$. Then the external angle $\gamma(G, \text{SP})$ between a $(l-1)$ -dimensional face G ($F \subseteq G$) and the skewed cross-polytope SP is given by

$$\gamma(G, \text{SP}) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx. \quad (79)$$

Proof: Without loss of generality, assume $K = \{n-k+1, \dots, n\}$. Consider the $(l-1)$ -dimensional face

$$G = \text{conv}\{C \times e^{n-l+1}, \dots, C \times e^{n-k}, e^{n-k+1}, \dots, e^n\}$$

of the skewed cross-polytope SP . The 2^{n-l} outward normal vectors of the supporting hyperplanes of the facets containing G are given by

$$\left\{ \sum_{p=1}^{n-l} j_p e_p / C + \sum_{p=n-l+1}^{n-k} e_p / C + \sum_{p=n-k+1}^n e_p, j_p \in \{-1, 1\} \right\}.$$

Then the outward normal cone $c(G, \text{SP})$ at the face G is the positive hull of these normal vectors. Thus

$$\begin{aligned} \int_{c(G, \text{SP})} e^{-\|x\|^2} dx &= \gamma(G, \text{SP}) V_{n-l}(S^{n-l}) \\ &\times \int_0^\infty e^{-r^2} r^{n-l} dr = \gamma(G, \text{SP}) \cdot \pi^{(n-l+1)/2}, \end{aligned} \quad (80)$$

where $V_{n-l}(S^{n-l})$ is the spherical volume of the $(n-l)$ -dimensional sphere S^{n-l} . Now define U to be the set

$$\{x \in \mathbb{R}^{n-l+1} \mid x_{n-l+1} \geq 0, |x_p| \leq \frac{x_{n-l+1}}{C}, 1 \leq p \leq (n-l)\}$$

and define $f(x_1, \dots, x_{n-l+1}) : U \rightarrow c(G, \text{SP})$ to be the linear and bijective map

$$\begin{aligned} f(x_1, \dots, x_{n-l+1}) &= \sum_{p=1}^{n-l} x_p e_p + \sum_{p=n-l+1}^{n-k} \frac{x_{n-l+1}}{C} e_p \\ &+ \sum_{p=n-k+1}^n x_{n-l+1} \times e_p. \end{aligned}$$

Then

$$\begin{aligned}
& \int_{c(G, \text{SP})} e^{-\|x'\|^2} dx' \\
&= \sqrt{k + \frac{l-k}{C^2}} \int_U e^{-\|f(x)\|^2} dx \\
&= \sqrt{k + \frac{l-k}{C^2}} \int_0^\infty \int_{-\frac{x_{n-l+1}}{C}}^{\frac{x_{n-l+1}}{C}} \cdots \int_{-\frac{x_{n-l+1}}{C}}^{\frac{x_{n-l+1}}{C}} \\
&\quad e^{-x_1^2 - \cdots - x_{n-l}^2 - (k + \frac{l-k}{C^2})x_{n-l+1}^2} dx_1 \cdots dx_{n-l+1} \\
&= \sqrt{k + \frac{l-k}{C^2}} \int_0^\infty e^{-(k + \frac{l-k}{C^2})x^2} \\
&\quad \times \left(\int_{-\frac{x}{C}}^{\frac{x}{C}} e^{-y^2} dy \right)^{n-l} dx \\
&= 2^{n-l} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k + \frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx,
\end{aligned}$$

where $\sqrt{k + \frac{l-k}{C^2}}$ is due to the change of integral variables. Combining it with (80) leads to the desired result. ■

E. Proof of Lemma 3

Consider any fixed $\delta > 0$. First, we consider the internal angle exponent ψ_{int} , where we define $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}$. Then for this fixed δ ,

$$\frac{1 - \gamma'}{\gamma'} \geq \frac{\frac{C^2-1}{C^2}\rho\delta + \frac{\delta}{C^2}}{\rho\delta} - 1$$

uniformly over $\nu \in [\delta, 1]$.

Now if we take ρ small enough, $\frac{\frac{C^2-1}{C^2}\rho\delta + \frac{\delta}{C^2}}{\rho\delta}$ can be arbitrarily large. By the asymptotic expression (31), this leads to large enough internal decay exponent ψ_{int} . At the same time, the external angle exponent ψ_{ext} is lower-bounded by zero and the combinatorial exponent is upper-bounded by some finite number. Then if ρ is small enough, we will get the net exponent ψ_{net} to be negative uniformly over the range $\nu \in [\delta, 1]$.

F. Proof of Lemma 4

We will show that for fixed $C > 1$, with $\rho(\delta) = \frac{1}{C^2} \log(1/\delta)^{-(1+\eta)}$ and some $\delta_0 > 0$,

$$\psi_{net}(\nu; \rho(\delta), \delta) < -\delta, \quad \delta < \delta_0, \quad \nu \in [\delta, 1].$$

To this end, we need to get the asymptotic of $\psi_{int}(\nu)$, $\psi_{ext}(\nu)$ and $\psi_{com}(\nu)$ as $\delta \rightarrow 0$ and $\rho(\delta) = \log(1/\delta)^{-(1+\eta)}$.

With

$$H(\nu) + H(\rho\delta/\nu)\nu = H(\rho\delta) + H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta),$$

from its definition, $\psi_{net}(\nu; \rho, \delta)$ is equal to

$$H(\nu) - \psi_{ext}(\nu) - \xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta) + \rho\delta \log(2) + H(\rho\delta/\nu)\nu.$$

From the derivation (or the expression) of the external angle $\gamma(G, \text{SP})$ in this paper, $\gamma(G, \text{SP})$ is a decreasing function in C . So we can upper-bound $\gamma(G, \text{SP})$ uniformly in $\nu \in [\delta, 1]$, for any $C \geq 1$, by

$$\frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{\sqrt{l}}} e^{-y^2} dy \right)^{n-l} dx,$$

namely the expression for the external angle when $C = 1$.

Now define $\Omega(\nu) = H(\nu) - \psi_{ext}^{C=1}(\nu)$, where $\psi_{ext}^{C=1}(\nu)$ is the external exponent when $C = 1$. Then from the asymptotic formula (27), we have

$$H(\nu) - \psi_{ext}(\nu) \leq \Omega(\nu) \sim \frac{1}{2} \log(\log(\frac{1}{\nu}))\nu,$$

as $\nu \rightarrow 0$.

So $\psi_{net}(\nu; \rho, \delta)$ is no bigger than

$$\Omega(\nu) - \xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta) + \rho\delta \log(2) + H(\rho\delta/\nu)\nu.$$

From [Don06b], for a certain δ_1 , if $\delta < \delta_1$,

$$H(\rho\delta/\nu) \leq H(\rho)\delta + 2\rho(\nu - \delta),$$

so we have

$$\psi_{net}(\nu) \leq K(\nu; \rho, \delta) + [\rho\delta \log(2) + H(\rho)\delta] + 2\rho(\nu - \delta),$$

where $K(\nu; \rho, \delta) \doteq \Omega(\nu) - \xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta)$.

As we will show later in Lemma 17, $K(\nu; \rho, \delta)$ is a concave function in $\nu \in [\delta, 1]$ if $\delta < \delta_2$. Also, we will show that for $\delta < \delta_3$,

$$K'(\delta; \rho, \delta) \leq -\eta/4 \log(\log(\frac{1}{\delta})) \quad (81)$$

$$K(\delta; \rho, \delta) \leq -\delta \times \eta/4 \log(\log(\frac{1}{\delta})) \quad (82)$$

where $K'(\nu; \rho, \delta) \doteq \frac{\partial K(\nu; \rho, \delta)}{\partial \nu}$. So there exists a $\delta_4 > 0$ so that for any $\delta \in (0, \delta_4)$,

$$K'(\delta; \rho, \delta) < -2\rho.$$

Also there exists a $\delta_5 > 0$ so that for any $0 < \delta < \delta_5$,

$$K(\delta; \rho, \delta) + \rho\delta \log(2) + H(\rho)\delta < -\delta.$$

Then by the concavity of $K(\nu; \rho, \delta)$, if $\delta < \min(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)$,

$$\psi_{net}(\nu) \leq -\delta$$

uniformly over the interval $\nu \in [\delta, 1]$.

Now we need to prove (81) and (82). As computed in [Don06b],

$$\Omega'(\nu) \sim \frac{1}{2} \log(\log(\frac{1}{\nu})), \quad \nu \rightarrow 0,$$

and

$$\Omega(\nu) \sim \frac{1}{2} \log(\log(\frac{1}{\nu}))\nu, \quad \nu \rightarrow 0.$$

By (31), we know that as $\delta \rightarrow 0$, and with $\nu = \delta$,

$$\xi_{\gamma'}(y_{\gamma'}) \sim \frac{1}{2} \log(\frac{1}{C^2\rho}) \sim \frac{1}{2} \log(\log(\frac{1}{\delta}))(1 + \eta).$$

Hence for $\delta < \delta_6$,

$$\xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta) \geq (1 + \frac{\eta}{2})\Omega(\delta). \quad (83)$$

Following this, there exists a $\delta_7 > 0$ so that for $\delta < \delta_7$,

$$K(\delta; \rho, \delta) \leq \frac{\eta}{4} \log \log(\frac{1}{\delta})\delta.$$

Also, from the asymptotic of $\Omega'(\nu)$ and the asymptotic of the derivative of $\xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta)$ with respect to ν in the next Lemma 17, we can further have

$$K'(\delta; \rho, \delta) \leq -\eta/4 \log(\log(\frac{1}{\delta})).$$

Lemma 17: If ρ is small enough, $K(\nu; \rho, \delta)$ is concave as a function of ν .

Proof: We define

$$\Upsilon(\nu; \rho, \delta) = \xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta).$$

Since $K(\nu; \rho, \delta) = \Omega(\nu) - \Upsilon(\nu; \rho, \delta)$ and $\Omega(\nu)$ is a concave function in ν [Don06b], we only need to show that $\Upsilon(\nu; \rho, \delta)$ is a convex function in ν .

Recall that $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}$ and we first look at $\frac{\partial\gamma'}{\partial\nu}$:

$$\frac{\partial\gamma'}{\partial\nu} = -\frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}} \times \frac{1}{(C^2-1)\rho\delta + \nu}.$$

So

$$\begin{aligned} \frac{\partial\Upsilon(\nu; \rho, \delta)}{\partial\nu} &= \xi_{\gamma'}(y_{\gamma'}) + \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \frac{\partial\gamma'}{\partial\nu} \cdot (\nu - \rho\delta) \\ &= \xi_{\gamma'}(y_{\gamma'}) - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \gamma' \cdot \frac{1 - \frac{\rho\delta}{\nu}}{(C^2-1)\frac{\rho\delta}{\nu} + 1} \end{aligned}$$

If we define $\Xi = \frac{1 - \frac{\rho\delta}{\nu}}{(C^2-1)\frac{\rho\delta}{\nu} + 1}$ and $\Pi = \frac{1}{(C^2-1)\rho\delta + \nu}$, we can have

$$\begin{aligned} &\frac{\partial^2\Upsilon(\nu; \rho, \delta)}{\partial\nu^2} \\ &= \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \frac{\partial\gamma'}{\partial\nu} - \frac{\partial^2\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'^2} \cdot \frac{\partial\gamma'}{\partial\nu} \cdot \gamma' \cdot \Xi \\ &\quad - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \frac{\partial\gamma'}{\partial\nu} \cdot \Xi - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \frac{\partial\Xi}{\partial\nu} \cdot \gamma' \\ &= \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot (-\gamma') \cdot (1 - \Xi) \cdot \Pi \\ &\quad - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \gamma' \cdot \frac{\partial\Xi}{\partial\nu} + \frac{\partial^2\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'^2} \cdot \gamma'^2 \cdot \Xi \cdot \Pi \\ &= -\frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \gamma'^2 \cdot \Pi - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \gamma'^2 \cdot \Pi \\ &\quad + \frac{\partial^2\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'^2} \cdot \gamma'^2 \cdot \Xi \cdot \Pi \end{aligned}$$

It has been shown in [Don06b] that as $\gamma \rightarrow 0$,

$$\begin{aligned} \frac{\partial\xi_{\gamma}(y_{\gamma})}{\partial\gamma} &\sim -\frac{\gamma^{-1}}{2}, \\ \frac{\partial^2\xi_{\gamma}(y_{\gamma})}{\partial\gamma^2} &\sim \frac{\gamma^{-4}}{4}. \end{aligned}$$

So from the definition of γ' , there exists a small enough ρ_0 such that for any $\rho < \rho_0$,

$$\frac{\partial^2 \Upsilon(\nu; \rho, \delta)}{\partial \nu^2} > 0, \nu \in [\delta, 1], \quad (84)$$

which then implies the concavity of $K(\nu; \rho, \delta)$. ■

G. Proof of Lemma 5

Proof: Suppose instead that $\rho_N(\delta, C) > \frac{1}{C+1}$. Then for every vector \mathbf{w} from the null space of the measurement matrix A , any $\rho_N(\delta, C)$ fraction of the n components in \mathbf{w} take no more than $\frac{1}{C+1}$ fraction of $\|\mathbf{w}\|_1$. But this can not be true if we consider the $\rho_N(\delta, C)$ fraction of \mathbf{w} with the largest magnitudes.

Now we only need to prove the lower bound for $\rho_N(\delta, C)$; in fact, we argue that

$$\rho_N(\delta, C) \geq \frac{\rho_N(\delta, C=1)}{C^2}.$$

We know from Lemma 3 that $\rho_N(\delta, C) > 0$ for any $C \geq 1$. Denote $\psi_{net}(C)$, $\psi_{com}(\nu; \rho, \delta, C)$, $\psi_{int}(\nu; \rho, \delta, C)$ and $\psi_{ext}(\nu; \rho, \delta, C)$ as the respective exponents for a certain C . Because $\rho_N(\delta, C=1) > 0$, for any $\rho = \rho_N(\delta, C=1) - \epsilon$, where $\epsilon > 0$ is an arbitrarily small number, the net exponent $\psi_{net}(C=1)$ is negative uniformly over $\nu \in [\delta, 1]$.

By examining the formula (18) for the external angle $\gamma(G, \text{SP})$, where G is a $(l-1)$ -dimensional face of the skewed cross-polytope SP , we have $\gamma(G, \text{SP})$ is a decreasing function in both k and C for a fixed l . So $\gamma(G, \text{SP})$ is upper-bounded by

$$\frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{\sqrt{l}}} e^{-y^2} dy \right)^{n-l} dx, \quad (85)$$

namely the expression for the external angle when $C=1$. Then for any $C > 1$ and any k , $\psi_{ext}(\nu; \rho, \delta, C)$ is lower-bounded by $\psi_{ext}(\nu; \rho, \delta, C=1)$.

Now let us check $\psi_{int}(\nu; \rho, \delta, C)$ by using the formula (29). With

$$\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}},$$

we have

$$\frac{1-\gamma'}{\gamma'} = -\frac{1}{C^2} + \frac{\nu}{C^2\rho\delta}. \quad (86)$$

Then for any fixed $\delta > 0$, if we take $\rho = \frac{\rho_N(\delta, C=1) - \epsilon}{C^2}$, where ϵ is an arbitrarily small positive number, then for any $\nu \geq \delta$, $\frac{1-\gamma'}{\gamma'}$ is an increasing function in C . So, following easily from its

definition, $\xi_{\gamma'}(y_{\gamma'})$ is an increasing function in C . This further implies that $\psi_{int}(\nu; \rho, \delta)$ is an increasing function in C if we take $\rho = \frac{\rho_N(\delta, C=1) - \epsilon}{C^2}$, for any $\nu \geq \delta$.

Also, for any fixed ν and δ , it is not hard to show that $\psi_{com}(\nu; \rho, \delta, C)$ is a decreasing function in C if $\rho = \frac{\rho_N(\delta, C=1)}{C^2}$. This is because in (14),

$$\binom{n}{k} \binom{n-k}{l-k} = \binom{n}{l} \binom{l}{k}.$$

Thus for any $C > 1$, if $\rho = \frac{\rho_N(\delta, C=1) - \epsilon}{C^2}$, the net exponent $\psi_{net}(C)$ is also negative uniformly over $\nu \in [\delta, 1]$. Since the parameter ϵ can be arbitrarily small, our claim and Lemma 5 then follow. ■

H. Proof of Lemma 13

Proof: First, we notice that for any $C > 1$,

$$\zeta_W(\delta) \geq \rho_N(\delta, C)\delta,$$

$$\zeta_{Sec}(\delta) \geq \rho_N(\delta, C)\delta,$$

$$\zeta_S(\delta) = \rho_N(\delta, C)\delta,$$

so by Lemma 3,

$$\zeta_W(\delta) > 0,$$

$$\zeta_{Sec}(\delta) > 0,$$

$$\zeta_S(\delta) > 0.$$

Now we will prove

$$\lim_{\delta \rightarrow 1} \zeta_W(\delta) = \delta.$$

As discussed in previous sections, we know that the decay exponent for the probability that the condition (56) is violated is equal to

$$H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta) - \xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta) - \psi_{ext}(\nu).$$

But from the derivations of the exponents, we know that

$$0 = \lim_{\delta \rightarrow 1} \sup_{\nu \in [\delta, 1]} \psi_{ext}(\nu),$$

$$0 = \lim_{\delta \rightarrow 1} \sup_{\nu \in [\delta, 1]} H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right),$$

$$\xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta) \geq \xi_{\gamma'(\delta)}(y_{\gamma'(\delta)}) (1 - \rho)\delta > 0, \quad \nu \geq \delta,$$

where

$$\gamma'(\delta) = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\delta}{C^2}} = \frac{\rho}{\frac{C^2-1}{C^2}\rho + \frac{1}{C^2}}.$$

Noticing that $\xi_{\gamma'(\delta)}(y_{\gamma'(\delta)}) (1 - \rho) > 0$ is only determined by ρ , for any $0 < \rho < 1$, there exists a big enough $\delta < 1$, such that

$$H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right) (1 - \rho\delta) - \xi_{\gamma'}(y_{\gamma'}) (\nu - \rho\delta) - \psi_{ext}(\nu) < 0.$$

uniformly over $[\delta, 1]$. Then it follows that

$$\lim_{\delta \rightarrow 1} \zeta_W(\delta) = 1,$$

for any $C > 1$. ■

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under grant no. CCF-0729203, by the David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking.

REFERENCES

- [AS92] Fernando Affentranger and Rolf Schneider. Random projections of regular simplices. *Discrete Comput. Geom.*, 7(3):219–226, 1992.
- [BCT09] J. D. Blanchard, C. Cartis, and J. Tanner. The restricted isometry property and ℓ_q regularization: Phase transitions for sparse approximation. 2009. Submitted for publication.
- [BDDW08] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.
- [Boo86] W. M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry*. Springer-Verlag, 1986. 2nd ed. San Diego, CA: Academic.
- [Can06] E. J. Candès. Compressive sampling. In *International Congress of Mathematicians. Vol. III*, pages 1433–1452. Eur. Math. Soc., Zürich, 2006.
- [CDD08] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k -term approximation. *J. Amer. Math. Soc.* 22 (2009), 211–231, 2008.
- [CM73] J. F. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38(5):826–844, 1973.

- [CRT05] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications of Pure and Applied Mathematics*, 59:1207–1223, 2005.
- [CRT06] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [CT05] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.
- [DMM10] D. Donoho, A. Maleki, and A. Montanari. The noise-sensitivity phase transition in compressed sensing. *arXiv:1004.1218*, 2010.
- [Don06a] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [Don06b] David Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete and Computational Geometry*, 35(4):617–652, 2006.
- [DT05] David L. Donoho and Jared Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA*, 102(27):9452–9457, 2005.
- [FN03] A. Feuer and A. Nemirovski. On sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 49(6):1579–1581, 2003.
- [Gem80] Stuart Geman. A limit theorem for the norm of random matrices. *Annals of Probability*, 8(2):252–261, 1980.
- [Grü68] Branko Grünbaum. Grassmann angles of convex polytopes. *Acta Math.*, 121:293–302, 1968.
- [Grü03] Branko Grünbaum. *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, Second Edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler.
- [Had79] H. Hadwiger. Gitterpunktanzahl im simplex und willssche vermutung. *Math. Ann.*, 239:271–288, 1979.
- [HN] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Trans. on Information Theory*, 52.
- [KBH99] Jr. Károly Böröczky and Martin Henk. Random projections of regular polytopes. *Arch. Math. (Basel)*, 73(6):465–473, 1999.
- [KT07] Boris S. Kashin and Vladimir N. Temlyakov. A remark on compressed sensing. *Mathematical Notes*, 82(5):748–755, November 2007.
- [KXAH09] M. Amin Khajehnejad, Weiyu Xu, Amir Salman Avestimehr, and Babak Hassibi. Weighted ℓ_1 minimization for sparse recovery with prior information. In *Proceedings of the International Symposium on Information Theory*, 2009.
- [LN06] N. Linial and I. Novik. How neighborly can a centrally symmetric polytope be? *Discrete and Computational Geometry*, 36(6):273–281, 2006.
- [McM75] Peter McMullen. Non-linear angle-sum relations for polyhedral cones and polytopes. *Math. Proc. Cambridge Philos. Soc.*, 78(2):247–261, 1975.
- [MP67] V. A. Marčenko and L. A. Pastur. Distributions of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik*, 1:457–483, 1967.
- [Mui05] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley-Interscience, 2005. 2nd ed.
- [RIC] <http://www.dsp.ece.rice.edu/cs/>.
- [RV] Mark Rudelson and Roman Vershynin. Geometric approach to error correcting codes and reconstruction of signals. *International Mathematical Research Notices*, 64.

- [San52] L. A. Santaló. Geometría integral en espacios de curvatura constante. *Rep.Argetina Publ.Com.Nac.Energí Atómica, Ser.Mat 1, No.1*, 1952.
- [Sil85] J. W. Silverstein. The smallest eigenvalue of a large dimensional wishart matrix. *The Annals of Probability*, 13:1364–1368, 1985.
- [Sto09] Mihailo Stojnic. Various thresholds for ℓ_1 -optimization in compressed sensing. 2009. Preprint available at <http://arxiv.org/abs/0907.3666>.
- [SXH08a] Mihailo Stojnic, Weiyu Xu, and Babak Hassibi. Compressed sensing - probabilistic analysis of a null-space characterization. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.
- [SXH08b] Mihailo Stojnic, Weiyu Xu, and Babak Hassibi. Compressed sensing of approximately sparse signals. In *IEEE International Symposium on Information Theory*, 2008.
- [TWD⁺06] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk. Random filters for compressive sampling and reconstruction. in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2006.
- [Vav09] S. Vavasis. Derivation of compressive sensing theorems for the spherical section property. *University of Waterloo, CO 769 lecture notes*, 2009.
- [VS92] A. M. Vershik and P. V. Sporyshev. Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Selecta Mathematica Sovietica*, 11(2):181–201, 1992.
- [Wai06] M. Wainwright. Sharp thresholds for noisy and high-dimensional recovery of sparsity using ℓ_1 -constrained quadratic programming. *Technical report, UC Berkeley, Department of Statistics*, 2006.
- [XH08] Weiyu Xu and Babak Hassibi. Compressed sensing over the grassmann manifold: A unified analytical framework. *Proceedings of the Forty-Sixth Annual Allerton Conference on Communication, Control, and Computing*, 2008.
- [XKAH10] Weiyu Xu, Amin Khajehnejad, Amir Salman Avestimehr, and Babak Hassibi. Breaking through the thresholds: an analysis for iterative reweighted ℓ_1 minimization via the Grassmann angle framework. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.
- [Zha06] Y. Zhang. When is missing data recoverable? 2006. available online at <http://www.caam.rice.edu/~zhang/reports/index.html>.
- [Zha08] Y. Zhang. Theory of compressive sensing via ℓ_1 -minimization: a non-rip analysis and extensions. 2008. available online at <http://www.caam.rice.edu/~zhang/reports/index.html>.